

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2003, 14:05:35 ; Search time 17.5 Seconds

(Without alignments)
98.565 Million cell updates/sec

Title: US-09-830-839-6

Perfect score: 77
Sequence: 1 WNFAGIEAASAI0G 15Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 440863 seqs, 114992915 residues

Total number of hits satisfying chosen parameters: 440863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubppaa/US08_NEM_PUB.pep:*
- 2: /cgn2_6/ptodata/1/pubppaa/PCT_NEM_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubppaa/US06_NEM_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubppaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubppaa/US07_NEM_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubppaa/PCTUS_PUBCOMB.pep:*
- 8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubppaa/US09_NEM_PUB.pep:*
- 10: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubppaa/US10_NEM_PUB.pep:*
- 12: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
- 13: /cgn2_6/ptodata/1/pubppaa/US60_NEM_PUB.pep:*
- 14: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|------------------------|--------------------|
| 1 | 77 | 100.0 | 20 | 9 US-10-044-703-60 | Sequence 60, Appl |
| 2 | 77 | 100.0 | 20 | 10 US-09-813-333-60 | Sequence 1, Appl |
| 3 | 77 | 100.0 | 95 | 9 US-08-805-427A-1 | Sequence 4, Appl |
| 4 | 77 | 100.0 | 403 | 9 US-09-805-427A-4 | Sequence 173, Appl |
| 5 | 77 | 100.0 | 404 | 9 US-09-805-427A-3 | Sequence 3, Appl |
| 6 | 77 | 100.0 | 404 | 10 US-09-791-171-173 | Sequence 12653, A |
| 7 | 77 | 100.0 | 404 | 10 US-09-791-171-172 | Sequence 1367, Ap |
| 8 | 66 | 85.7 | 20 | 9 US-10-044-703-61 | Sequence 61, Appl |
| 9 | 66 | 85.7 | 20 | 10 US-09-813-333-61 | Sequence 59, Appl |
| 10 | 60 | 77.9 | 13 | 9 US-10-044-703-59 | Sequence 59, Appl |
| 11 | 60 | 77.9 | 13 | 10 US-09-813-333-59 | Sequence 59, Appl |
| 12 | 41 | 53.2 | 463 | 9 US-10-156-761-12653 | Sequence 155, App |
| 13 | 39 | 50.6 | 9 | 10 US-09-916-201-13 | Sequence 155, App |
| 14 | 39 | 50.6 | 221 | 9 US-10-054-988-155 | Sequence 155, App |
| 15 | 39 | 50.6 | 221 | 10 US-09-739-254-155 | Sequence 155, App |
| 16 | 39 | 50.6 | 221 | 10 US-09-304-615-155 | Sequence 155, App |
| 17 | 39 | 50.6 | 239 | 9 US-10-103-313-351 | Sequence 351, App |
| 18 | 38 | 49.4 | 37 | 10 US-09-864-761-39026 | Sequence 39026, A |
| 19 | 37 | 48.1 | 153 | 10 US-09-925-301-1367 | Sequence 1367, Ap |

| | | | | | |
|----|------|------|-----|-----------------------|-------------------|
| 20 | 37 | 48.1 | 360 | 9 US-10-024-828-9 | Sequence 9, Appl |
| 21 | 37 | 47.4 | 739 | 9 US-10-156-761-10111 | Sequence 10111, A |
| 22 | 36.5 | 47.4 | 516 | 9 US-10-156-761-13151 | Sequence 13151, A |
| 23 | 36 | 46.8 | 98 | 10 US-09-894-018-101 | Sequence 101, App |
| 24 | 36 | 46.8 | 140 | 9 US-09-341-894-2 | Sequence 2, Appl |
| 25 | 36 | 46.8 | 254 | 9 US-09-902-525-30 | Sequence 30, Appl |
| 26 | 36 | 46.8 | 308 | 10 US-09-894-018-105 | Sequence 105, App |
| 27 | 36 | 46.8 | 308 | 10 US-09-894-018-107 | Sequence 107, App |
| 28 | 36 | 46.8 | 376 | 9 US-10-224-567-1 | Sequence 1, Appl |
| 29 | 35.5 | 46.1 | 679 | 10 US-09-815-242-5082 | Sequence 5082, Ap |
| 30 | 35 | 45.5 | 90 | 9 US-09-989-920-262 | Sequence 262, App |
| 31 | 35 | 45.5 | 159 | 9 US-09-882-691-18 | Sequence 18, Appl |
| 32 | 35 | 45.5 | 165 | 9 US-09-925-299-865 | Sequence 865, App |
| 33 | 35 | 45.5 | 165 | 10 US-09-925-299-865 | Sequence 865, App |
| 34 | 35 | 45.5 | 226 | 9 US-09-867-159A-2 | Sequence 2, Appl |
| 35 | 35 | 45.5 | 296 | 9 US-10-128-714-3095 | Sequence 3095, Ap |
| 36 | 35 | 45.5 | 296 | 9 US-10-128-714-8095 | Sequence 2, Appl |
| 37 | 35 | 45.5 | 320 | 9 US-09-877-160-2 | Sequence 79, Appl |
| 38 | 35 | 45.5 | 320 | 9 US-09-847-208-79 | Sequence 73, Appl |
| 39 | 35 | 45.5 | 321 | 9 US-09-847-208-73 | Sequence 3900, Ap |
| 40 | 35 | 45.5 | 331 | 9 US-09-738-626-3900 | Sequence 290, App |
| 41 | 35 | 45.5 | 373 | 9 US-09-934-435-290 | Sequence 320, App |
| 42 | 35 | 45.5 | 482 | 9 US-10-153-668-320 | Sequence 432, App |
| 43 | 35 | 45.5 | 486 | 9 US-10-153-668-432 | Sequence 170, App |
| 44 | 35 | 45.5 | 489 | 9 US-09-934-455-170 | Sequence 3710, Ap |
| 45 | 35 | 45.5 | 495 | 9 US-09-738-626-3710 | |

ALIGNMENTS

RESULT 1
US-10-044-703-60
; Sequence 60, Application US/10044703
; Publication No. US20020192233A1
; GENERAL INFORMATION:
; APPLICANT: Degroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/10/044,703
; PRIOR FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 60
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-044-703-60

Query Match 100.0%; Score 77; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.6e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WNFAGIEAASAI0G 15
|||||
Db 2 WNFAGIEAASAI0G 16

RESULT 2
US-09-813-333-60
; Sequence 60, Application US/09813333
; Patent No. US20020119160A1
; GENERAL INFORMATION:
; APPLICANT: Degroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/09/813,333
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81

```
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 60
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-813-333-60
```

```
Query Match          100.0%; Score 77; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.6e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

OY 1 WNFAGIEAASAIOG 15
|||||

DB 2 WNFAGIEAASAIOG 16

```
RESULT 3
US-09-805-427A-1
; Sequence 1, Application US/09805427A
; Patent No. US20020176867A1
; GENERAL INFORMATION:
; APPLICANT: Statens Serum Institut
; TITLE OF INVENTION: Hybrids of M. tuberculosis Antigens
; FILE REFERENCE: 670001-2002.5
; CURRENT APPLICATION NUMBER: US/09/805,427A
; CURRENT FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 1
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-805-427A-1
```

```
Query Match          100.0%; Score 77; DB 9; Length 95;
Best Local Similarity 100.0%; Pred. No. 8.6e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

OY 1 WNFAGIEAASAIOG 15
|||||

DB 6 WNFAGIEAASAIOG 20

```
RESULT 4
US-09-805-427A-4
; Sequence 4, Application US/09805427A
; Patent No. US20020176867A1
; GENERAL INFORMATION:
; APPLICANT: Statens Serum Institut
; TITLE OF INVENTION: Hybrids of M. tuberculosis Antigens
; FILE REFERENCE: 670001-2002.5
; CURRENT APPLICATION NUMBER: US/09/805,427A
; CURRENT FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 4
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Fusion protein ESAT-6-Ag85B
US-09-805-427A-4
```

```
Query Match          100.0%; Score 77; DB 9; Length 403;
Best Local Similarity 100.0%; Pred. No. 4e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

OY 1 WNFAGIEAASAIOG 15
|||||

DB 27 WNFAGIEAASAIOG 41

RESULT 5

```
US-09-791-171-173
; Sequence 173, Application US/09791171
; Patent No. US20020094336A1
; GENERAL INFORMATION:
; APPLICANT: ANDERSEN, Peter
; APPLICANT: NIELSEN, Rikke
; APPLICANT: OETTINGER, Thomas
; APPLICANT: RASMUSSEN, Peter Blak
; APPLICANT: ROSENKRANDS, Ida
; APPLICANT: WELDRING, Karin
; APPLICANT: FLORIO, Walter
; TITLE OF INVENTION: NUCLEIC ACIDS FRAGMENTS AND POLYPEPTIDE FRAGMENTS
; TITLE OF INVENTION: DERIVED FROM M. TUBERCULOSIS
; FILE REFERENCE: 670001-2002.1
; CURRENT APPLICATION NUMBER: US/09/791,171
; CURRENT FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 09/050,739
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 0376/97
; PRIOR FILING DATE: 1997-04-02
; PRIOR APPLICATION NUMBER: 1277/97
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/044,624
; PRIOR FILING DATE: 1997-04-18
; PRIOR APPLICATION NUMBER: 60/070,488
; PRIOR FILING DATE: 1998-01-05
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 173
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-791-171-173
```

```
Query Match          100.0%; Score 77; DB 10; Length 403;
Best Local Similarity 100.0%; Pred. No. 4e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

OY 1 WNFAGIEAASAIOG 15
|||||

DB 27 WNFAGIEAASAIOG 41

```
RESULT 6
US-09-805-427A-3
; Sequence 3, Application US/09805427A
; Patent No. US20020176867A1
; GENERAL INFORMATION:
; APPLICANT: Statens Serum Institut
; TITLE OF INVENTION: Hybrids of M. tuberculosis Antigens
; FILE REFERENCE: 670001-2002.5
; CURRENT APPLICATION NUMBER: US/09/805,427A
; CURRENT FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 3
; LENGTH: 404
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Fusion protein Ag85B-ESAT-6
US-09-805-427A-3
```

```
Query Match          100.0%; Score 77; DB 9; Length 404;
Best Local Similarity 100.0%; Pred. No. 4e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

OY 1 WNFAGIEAASAIOG 15
|||||

DB 315 WNFAGIEAASAIOG 329

RESULT 7

US-09-791-171-172
; Sequence 172, Application US/09791171
; Patent No. US20020094336A1
; GENERAL INFORMATION:
; APPLICANT: ANDERSEN, Peter
; APPLICANT: NIELSEN, Rikke
; APPLICANT: OETTINGER, Thomas
; APPLICANT: RASMUSSEN, Peter Birk
; APPLICANT: ROSENKRANDS, Ida
; APPLICANT: WEIDINGH, Karin
; APPLICANT: FLORIO, Walter
; TITLE OF INVENTION: NUCLEIC ACIDS FRAGMENTS AND POLYPEPTIDE FRAGMENTS
; FILE REFERENCE: DERIVED FROM M. TUBERCULOSIS
; CURRENT APPLICATION NUMBER: US/09/791,171
; CURRENT FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 09/050,739
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 0376/97
; PRIOR FILING DATE: 1997-04-02
; PRIOR APPLICATION NUMBER: 1277/97
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/044,624
; PRIOR FILING DATE: 1997-04-18
; PRIOR APPLICATION NUMBER: 60/070,488
; PRIOR FILING DATE: 1998-01-05
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 172
; LENGTH: 404
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-791-171-172

Query Match 100.0%; Score 77; DB 10; Length 404;
Best Local Similarity 100.0%; Pred. No. 4e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NFAGIEAASAIQ 15
DB 315 WNFAGIEAASAIQ 329

RESULT 8
US-10-044-703-61
; Sequence 61, Application US/10044703
; Publication No. US20020192233A1
; GENERAL INFORMATION:
; APPLICANT: Degroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/10/044,703
; CURRENT FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 61
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-044-703-61

Query Match 85.7%; Score 66; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.00012;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 NFAGIEAASAIQ 15
DB 1 NFAGIEAASAIQ 14

RESULT 9

US-09-813-333-61
; Sequence 61, Application US/09813333
; Patent No. US20020119160A1
; GENERAL INFORMATION:
; APPLICANT: Degroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/09/813,333
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 61
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-813-333-61

Query Match 85.7%; Score 66; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.00012;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 NFAGIEAASAIQ 15
DB 1 NFAGIEAASAIQ 14

RESULT 10
US-10-044-703-59
; Sequence 59, Application US/10044703
; Publication No. US20020192233A1
; GENERAL INFORMATION:
; APPLICANT: Degroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/10/044,703
; CURRENT FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 59
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-044-703-59

Query Match 77.9%; Score 60; DB 9; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.00076;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 NFAGIEAASAIQ 14
DB 1 NFAGIEAASAIQ 13

RESULT 11
US-09-813-333-59
; Sequence 59, Application US/09813333
; Patent No. US20020119160A1
; GENERAL INFORMATION:
; APPLICANT: Degroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/09/813,333
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 59
; LENGTH: 13

;; TYPE: PRT
;; ORGANISM: Mycobacterium tuberculosis
US-09-813-333-55

Query Match 77.9%; Score 60; DB 10; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.00076;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 NFAGIEAASAIQ 14
DB 1 NFAGIEAASAIQ 13

RESULT 12
US-10-156-761-12653
; Sequence 12653, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12653
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12653

Query Match 53.2%; Score 41; DB 9; Length 463;
Best Local Similarity 53.8%; Pred. No. 53;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 WNFAGIEAASAI 13
DB 311 WNFAGIEAASAI 323

RESULT 13
US-09-916-201-13
; Sequence 13, Application US/09916201
; Patent No. US20020131976A1
; GENERAL INFORMATION:
; APPLICANT: LAIVANT, AJIT
; APPLICANT: PRATHAN, ANSAT A.
; APPLICANT: HILL, ADRIAN V.S.
; TITLE OF INVENTION: TUBERCULOSIS VACCINE
; FILE REFERENCE: 117-359
; CURRENT APPLICATION NUMBER: US/09/916,201
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 09/467,893
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: US 60/113,783
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 9
; TYPE: PRT
; ORGANISM: M. tuberculosis
US-09-916-201-13

Query Match 50.6%; Score 39; DB 10; Length 9;
Best Local Similarity 100.0%; Pred. No. 3.9e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GIEAASAI 13
DB 1 GIEAASAI 9

RESULT 14
US-10-054-988-155
; Sequence 155, Application US/10054988
; Publication No. US20030087341A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 49 Human Secreted Proteins
; FILE REFERENCE: P2032P1
; CURRENT APPLICATION NUMBER: US/10/054,988
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: 09/904,615
; PRIOR FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: 09/511,554
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/097,917
; PRIOR FILING DATE: 1998-08-25
; PRIOR APPLICATION NUMBER: 60/098,634
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 155
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-054-988-155

Query Match 50.6%; Score 39; DB 9; Length 221;
Best Local Similarity 57.1%; Pred. No. 53;
Matches 8; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 WNFAGIEAASAIQ 14
DB 132 WNFAGIEAASAIQ 145

RESULT 15
US-09-739-254-155
; Sequence 155, Application US/09739254
; Patent No. US20010021700A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 49 Human Secreted Proteins
; FILE REFERENCE: P2032P1
; CURRENT APPLICATION NUMBER: US/09/739,254
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/511,554
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: PCT/US99/19330
; PRIOR FILING DATE: 1999-08-24
; PRIOR APPLICATION NUMBER: 60/097,917
; PRIOR FILING DATE: 1998-08-25
; PRIOR APPLICATION NUMBER: 60/098,634
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 155
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-739-254-155

Query Match 50.6%; Score 39; DB 10; Length 221;
Best Local Similarity 57.1%; Pred. No. 53;
Matches 8; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Thu Jul 3 14:50:03 2003

us-09-830-839-6.rapb

Page 5

OY 1 WNFAGIEAASATQ 14
|| | ||| |
Db 132 WNLGDEAAALAQ 145

Search completed: July 3, 2003, 14:08:27
Job time : 17.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2003, 14:04:30 ; Search time 11 Seconds
(without alignments)
40.122 Million cell updates/sec

Title: US-09-830-839-6
Perfect score: 77
Sequence: 1 WNFAGIEAASAIQG 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Issued Patents_Aa.*

1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PC10S.COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------------|
| 1 | 77 | 100.0 | 51 | 4 | US-08-818-112-104 |
| 2 | 77 | 100.0 | 51 | 4 | US-08-818-111-99 |
| 3 | 77 | 100.0 | 51 | 4 | US-09-056-556-104 |
| 4 | 77 | 100.0 | 51 | 4 | US-09-072-596-99 |
| 5 | 77 | 100.0 | 95 | 2 | US-08-465-640-2 |
| 6 | 40 | 51.9 | 2293 | 4 | US-09-368-590-2 |
| 7 | 38 | 49.4 | 418 | 4 | US-09-202-893B-4 |
| 8 | 37 | 48.1 | 360 | 4 | US-09-509-902A-9 |
| 9 | 35 | 45.5 | 29 | 2 | US-08-482-142-11 |
| 10 | 35 | 45.5 | 29 | 2 | US-08-482-142-74 |
| 11 | 35 | 45.5 | 29 | 2 | US-08-478-572-11 |
| 12 | 35 | 45.5 | 29 | 2 | US-08-478-572-74 |
| 13 | 35 | 45.5 | 29 | 4 | US-08-484-296-11 |
| 14 | 35 | 45.5 | 29 | 4 | US-08-484-296-74 |
| 15 | 35 | 45.5 | 29 | 5 | US-08-484-296-11 |
| 16 | 35 | 45.5 | 117 | 4 | US-09-042-353-232 |
| 17 | 35 | 45.5 | 117 | 4 | US-08-758-417A-80 |
| 18 | 35 | 45.5 | 159 | 4 | US-09-347-803-18 |
| 19 | 35 | 45.5 | 222 | 1 | US-07-945-288-11 |
| 20 | 35 | 45.5 | 222 | 1 | US-08-462-831-11 |
| 21 | 35 | 45.5 | 222 | 1 | US-08-461-809-11 |
| 22 | 35 | 45.5 | 222 | 1 | US-08-461-809-11 |
| 23 | 35 | 45.5 | 222 | 1 | US-08-461-809-11 |
| 24 | 35 | 45.5 | 245 | 1 | US-07-945-288-2 |
| 25 | 35 | 45.5 | 245 | 1 | US-08-462-831-2 |
| 26 | 35 | 45.5 | 245 | 1 | US-08-461-809-2 |
| 27 | 35 | 45.5 | 245 | 1 | US-08-461-809-2 |

| | | | | | | |
|----|----|------|-----|---|-------------------|--------------------|
| 28 | 35 | 45.5 | 245 | 2 | US-08-482-142-2 | Sequence 2, Appl1 |
| 29 | 35 | 45.5 | 245 | 2 | US-08-478-572-2 | Sequence 2, Appl1 |
| 30 | 35 | 45.5 | 245 | 3 | US-08-460-040-2 | Sequence 2, Appl1 |
| 31 | 35 | 45.5 | 245 | 4 | US-08-484-296-2 | Sequence 2, Appl1 |
| 32 | 35 | 45.5 | 245 | 5 | PCT-US93-08518-2 | Sequence 2, Appl1 |
| 33 | 35 | 45.5 | 320 | 1 | US-07-945-288-10 | Sequence 10, Appl1 |
| 34 | 35 | 45.5 | 320 | 1 | US-08-462-831-10 | Sequence 10, Appl1 |
| 35 | 35 | 45.5 | 320 | 1 | US-08-461-809-10 | Sequence 10, Appl1 |
| 36 | 35 | 45.5 | 320 | 1 | US-08-482-142-6 | Sequence 10, Appl1 |
| 37 | 35 | 45.5 | 320 | 1 | US-08-461-809-10 | Sequence 10, Appl1 |
| 38 | 35 | 45.5 | 321 | 1 | PCT-US93-08518-10 | Sequence 10, Appl1 |
| 39 | 35 | 45.5 | 321 | 1 | US-07-945-288-6 | Sequence 10, Appl1 |
| 40 | 35 | 45.5 | 321 | 1 | US-08-462-831-6 | Sequence 6, Appl1 |
| 41 | 35 | 45.5 | 321 | 1 | US-08-461-809-6 | Sequence 6, Appl1 |
| 42 | 35 | 45.5 | 321 | 2 | US-08-482-142-6 | Sequence 6, Appl1 |
| 43 | 35 | 45.5 | 321 | 2 | US-08-478-572-6 | Sequence 6, Appl1 |
| 44 | 35 | 45.5 | 321 | 4 | US-08-484-296-6 | Sequence 6, Appl1 |
| 45 | 35 | 45.5 | 321 | 5 | PCT-US93-08518-6 | Sequence 6, Appl1 |

ALIGNMENTS

RESULT 1
US-08-818-112-104
Sequence 104, Application US/08818112

Patent No. 6290969

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Dillon, Davin C.

APPLICANT: Campos-Neto, Antonio

APPLICANT: Houghton, Raymond

APPLICANT: Vedicik, Thomas S.

APPLICANT: Twardzik, Daniel R.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY

TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 153

CORRESPONDENCE ADDRESS:

ADDRESSER: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/818,112

FILING DATE: 13-MAR-1997

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Makl, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.411C6

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 104:

SEQUENCE CHARACTERISTICS:

LENGTH: 51 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-818-112-104

Query Match

Best Local Similarity 100.0%; Score 77; DB 4; Length 51;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WNFAGIEAASAIQG 15
Db 6 WNFAGIEAASAIQG 20

RESULT 2

US-08-818-111-99
Sequence 99, Application US/08818111
Patent No. 6338852
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Vardzik, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 148
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,111
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 99:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-111-99

Query Match 100.0%; Score 77; DB 4; Length 51;
Best Local Similarity 100.0%; Pred. No. 7.6e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WNFAGIEAASAIQG 15
Db 6 WNFAGIEAASAIQG 20

RESULT 3

US-09-056-556-104
Sequence 104, Application US/09056556
Patent No. 6350456
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle

TREATY

STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-056-556-104

Query Match 100.0%; Score 77; DB 4; Length 51;
Best Local Similarity 100.0%; Pred. No. 7.6e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WNFAGIEAASAIQG 15
Db 6 WNFAGIEAASAIQG 20

RESULT 4

US-09-072-596-99
Sequence 99, Application US/09072596
Patent No. 6458366
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Vardzik, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 99:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-072-596-99

Query Match 100.0%; Score 77; DB 4; Length 51;
Best Local Similarity 100.0%; Pred. No. 7.6e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WNFAGIEAASAI0G 15
DB 6 WNFAGIEAASAI0G 20

RESULT 5

US-08-465-640-2
Sequence 2, Application US/08465640
Patent No. 5953077

GENERAL INFORMATION:
APPLICANT: ANDERSEN, Peter
APPLICANT: ANDERSEN, Ase Bengaard
APPLICANT: HASLOV, Kaare
APPLICANT: SORENSEN, Anne Lund
TITLE OF INVENTION: TUBERCULOSIS VACCINE
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,640
FILING DATE: 05-JUN-1995

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/123,182
FILING DATE: 20-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DK94/00273
FILING DATE: 01-JUL-1994

ATTORNEY/AGENT INFORMATION:
NAME: COOPER, IVER P
REGISTRATION NUMBER: 28,005
REFERENCE/DOCKET NUMBER: ANDERSEN-3A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 95 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-465-640-2

Query Match 100.0%; Score 77; DB 2; Length 95;
Best Local Similarity 100.0%; Pred. No. 1.6e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WNFAGIEAASAI0G 15
DB 6 WNFAGIEAASAI0G 20

DB 6 WNFAGIEAASAI0G 20

RESULT 6
US-09-368-590-2
Sequence 2, Application US/09368590
Patent No. 6187563

GENERAL INFORMATION:
APPLICANT: Solimena, Michele
TITLE OF INVENTION: INTERACTING POLYPEPTIDES FOR
FILE REFERENCE: 101918-200 (OCR-941)
CURRENT APPLICATION NUMBER: US/09/368,590
CURRENT FILING DATE: 1999-08-04
EARLIER APPLICATION NUMBER: 60/095,657
EARLIER FILING DATE: 1998-08-07
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 2293
TYPE: PRT
ORGANISM: Human
US-09-368-590-2

Query Match 51.9%; Score 40; DB 4; Length 2293;
Best Local Similarity 42.9%; Pred. No. 2e+02;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 WNFAGIEAASAI0G 14
DB 720 WRLSGLEAALQALE 733

RESULT 7

US-09-202-893B-4
Sequence 4, Application US/09202893B
Patent No. 6319692

GENERAL INFORMATION:
APPLICANT: KADOTA, Mariko
APPLICANT: KIMAKI, Mayumi
APPLICANT: SAWAKI, Saeko
APPLICANT: SHIRASAWA, Yukiko
APPLICANT: SONE, Harue
APPLICANT: SAKO, Tomoyuki
TITLE OF INVENTION: METHODS FOR TRANSFERRING GENE INTO CHROMOSOME
FILE REFERENCE: 980794/HG
CURRENT APPLICATION NUMBER: US/09/202,893B
CURRENT FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: PCT/JP97/02187
PRIOR FILING DATE: 1997-06-25
PRIOR APPLICATION NUMBER: JP 8/184266
PRIOR FILING DATE: 1996-06-26
PRIOR APPLICATION NUMBER: JP 8/257764
PRIOR FILING DATE: 1996-09-06

NUMBER OF SEQ ID NOS: 9
SEQ ID NO 4
LENGTH: 418
TYPE: PRT
ORGANISM: Lactobacillus casei
US-09-202-893B-4

Query Match 49.4%; Score 38; DB 4; Length 418;
Best Local Similarity 61.5%; Pred. No. 62;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 FAGIEAASAI0G 15
DB 31 FARIKAAAKKY0G 43

RESULT 8
US-09-509-902A-9
Sequence 9, Application US/09509902A

Patent No. 6387676
GENERAL INFORMATION:
APPLICANT: Virca, Duke
APPLICANT: Bird, Timothy A.
APPLICANT: Anderson, Dirk M.
APPLICANT: Marken, John S.
TITLE OF INVENTION: Human CDNAs Encoding Polypeptides Having Kinase Functions
FILE REFERENCE: 2877-US
CURRENT APPLICATION NUMBER: US/09/509,902A
CURRENT FILING DATE: 1999-08-03
NUMBER OF SEQ ID NOS: 16
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 9
LENGTH: 360
TYPE: PRT
ORGANISM: Homo sapiens
US-09-509-902A-9

Query Match 48.1%; Score 37; DB 4; Length 360;
Best Local Similarity 50.0%; Pred. No. 78;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 1 WNFAGTEAASA 12
Db 4 WSWAGIPSSAAA 15

RESULT 9

US-08-482-142-11
Sequence 11, Application US/08482142
Patent No. 5820862
GENERAL INFORMATION:
APPLICANT: Garman, Richard
APPLICANT: Greenstein, Julia
APPLICANT: Kuo, Mei-chang
APPLICANT: Rogers, Bruce
APPLICANT: Franzen, Henry
APPLICANT: Chen, Xian
APPLICANT: Evans, Sean
APPLICANT: Shaked, Ze'ev
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
STREET: 610 LINCOLN STREET
CITY: WALTHAM
STATE: MA
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,142
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/445,307
FILING DATE: 07 June 1995
ATTORNEY/AGENT INFORMATION:
NAME: CRAIG, ANNE I.
REGISTRATION NUMBER: 32,976
REFERENCE/DOCKET NUMBER: 017,605
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 466-6000
TELEFAX: (617) 466-6040
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 29 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: N-terminal
US-08-482-142-11

Query Match 45.5%; Score 35; DB 2; Length 29;
Best Local Similarity 50.0%; Pred. No. 9.2;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 WNFAGTEAASA 12
Db 15 WAFSGVAATESA 26

RESULT 10

US-08-482-142-74
Sequence 74, Application US/08482142
Patent No. 5820862
GENERAL INFORMATION:
APPLICANT: Garman, Richard
APPLICANT: Greenstein, Julia
APPLICANT: Kuo, Mei-chang
APPLICANT: Rogers, Bruce
APPLICANT: Franzen, Henry
APPLICANT: Chen, Xian
APPLICANT: Evans, Sean
APPLICANT: Shaked, Ze'ev
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
STREET: 610 LINCOLN STREET
CITY: WALTHAM
STATE: MA
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,142
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/445,307
FILING DATE: 07 June 1995
ATTORNEY/AGENT INFORMATION:
NAME: CRAIG, ANNE I.
REGISTRATION NUMBER: 32,976
REFERENCE/DOCKET NUMBER: 017,605
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 466-6000
TELEFAX: (617) 466-6040
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 29 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: N-terminal
US-08-482-142-74

Query Match 45.5%; Score 35; DB 2; Length 29;
Best Local Similarity 50.0%; Pred. No. 9.2;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 WNFAGTEAASA 12
Db 15 WAFSGVAATESA 26

```

1  APPLICANT: Shaked, Ze'ev
2  TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
3  TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
4  NUMBER OF SEQUENCES: 207
5  CORRESPONDENCE ADDRESS:
6  ADDRESSEE: IMMUNOLOGIC PHARMACEUTICAL CORPORATION
7  STREET: 610 LINCOLN STREET
8  CITY: WALTHAM
9  STATE: MA
10 COUNTRY: USA
11 ZIP: 02154
12
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15
16 OPERATING SYSTEM: PC-DOS/MS-DOS
17 SOFTWARE: ASCII TEXT
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/08/478,572
20 FILING DATE: 07-June-1995
21 CLASSIFICATION:
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: 08/445,307
24 FILING DATE:
25 ATTORNEY/AGENT INFORMATION:
26 NAME: CRAIG, ANNE I.
27 REGISTRATION NUMBER: 32,976
28 REFERENCE/DOCKET NUMBER: 017,605
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE: (617) 466-6000
31 TELEFAX: (617) 466-6040
32 INFORMATION FOR SEQ ID NO: 74:
33 SEQUENCE CHARACTERISTICS:
34 LENGTH: 29 amino acids
35 TYPE: amino acid
36 TOPOLOGY: linear
37 MOLECULE TYPE: peptide
38 FRAGMENT TYPE: N-terminal
39
40 US-08-478-572-74
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
96
```

```

1      COMPUTER: IBM PC compatible
2      OPERATING SYSTEM: PC-DOS/MS-DOS
3      SOFTWARE: ASCII TEXT
4      CURRENT APPLICATION DATA:
5      APPLICATION NUMBER: US/08/484,296
6      FILING DATE:
7      CLASSIFICATION: 435
8      PRIOR APPLICATION DATA:
9      APPLICATION NUMBER: 08/445,307
10     FILING DATE: 07 June 1995
11     ATTORNEY/AGENT INFORMATION:
12     NAME: CRAIG, ANNE I.
13     REGISTRATION NUMBER: 32,976
14     REFERENCE/DOCKET NUMBER: 017,605
15     TELECOMMUNICATION INFORMATION:
16     TELEPHONE: (617) 466-6000
17     TELEFAX: (617) 466-6040
18     INFORMATION FOR SEQ ID NO: 11:
19     SEQUENCE CHARACTERISTICS:
20     LENGTH: 29 amino acids
21     TYPE: amino acid
22     TOPOLOGY: linear
23     MOLECULE TYPE: peptide
24     FRAGMENT TYPE: N-terminal
25     US-08-484-296-11

```

| | | | | |
|-----------------------|----------------|---------------|----------|-----------|
| Query Match | 45.5% | Score 35 | DB 4 | Length 29 |
| Best Local Similarity | 50.0% | Pred. NO. 9.2 | | |
| Matches 6 | Conservative 2 | Mismatches 4 | Indels 0 | Gaps 0 |

```
QY      1 WNFAGIEAASA 12
          | | : | : |
Db      15 WAFSGVAATESA 26
```

RESULT 14
US-08-484-296-74

Sequence 74, Application US/08484296
Patent No. 6268491

GENERAL INFORMATION:

APPLICANT: Garmen, Richard
APPLICANT: Greenstein, Julia
APPLICANT: Kuo, Mei-Chang
APPLICANT: Rogers, Bruce
APPLICANT: Franzen, Henry
APPLICANT: Chen, Xian
APPLICANT: Evans, Sean

TITLE OF INVENTION: T CELL EPTIPOES OF THE MAJOR ALLERGENS
TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
NUMBER OF SEQUENCES: 207

CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMUNOLOGIC PHARMACEUTICAL CORPORATION
STREET: 610 LINCOLN STREET
CITY: WALTHAM
STATE: MA
COUNTRY: USA
ZIP: 02154

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,296
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/445,307
FILING DATE: 07 June 1995

ATTORNEY/AGENT INFORMATION:
NAME: CRAIG, ANNE I.
REGISTRATION NUMBER: 32,976

```

: REFERENCE/DOCKET NUMBER: 017-6000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 466-6000
: TELEFAX: (617) 466-6040
: INFORMATION FOR SEQ ID NO: 74:
:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 29 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
:
: MOLECULE TYPE: peptide
: FRAGMENT TYPE: N-terminal
:
US-08-484-296-74

```

| | | | | |
|-----------------------|-------|--------------|------|--------------|
| Query Match | 45.5% | Score 35 | DB 4 | Length 29 |
| Best Local Similarity | 50.0% | Pred. No | 9.2 | |
| Matches | 6 | Conservative | 2 | Mismatches 4 |
| | | | | Indels 0 |
| | | | | Gaps 0 |

```
QY      1 WNFAGIEAASA 12
          | | : | |
Db      15 WAFSGVAATESA 26
```

RESULT 15
PCT-US95-04481-3

```

GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Pharmaceutical Peptide Formulations For Treatment of Dust M

```

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04481

```

CLASSIFICATION: .
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/227,772

; FILING DATE: April 14, 1994
 ;
 ATTORNEY/AGENT INFORMATION:
 ; NAME: Vanstone, Darlene A.
 ; REGISTRATION NUMBER: 35,279
 ;

TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 466-6000
TELEFAX: (617) 466-6040

```

; INFORMATION FOR SEQ ID NO: 3:
;
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 29 amino acids
;   TYPE: amino acid
;

```

```

; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal

```

| | | | | |
|-----------------------|-------|----------------|-------|-----------|
| Query Match | 45.5% | Score 35; | DB 5; | Length 29 |
| Best Local Similarity | 50.0% | Pred. No. 9.2; | | |

| | | | | | | | | | | |
|----|---------|-------------|--------------|----|------------|----|--------|----|------|----|
| | Matches | 6; | Conservative | 2; | Mismatches | 4; | Indels | 0; | Gaps | 0; |
| OY | 1 | WNFAGIEAASA | 12 | | | | | | | |
| | | 1 | 1 | : | : | : | : | : | : | : |

Search completed: July 3, 2003, 14:07:46
Job time : 11 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2003, 14:03:09 ; Search time 24 Seconds
(without alignments)
128.779 Million cell updates/sec

Title: US-09-830-839-6
Perfect score: 77
Sequence: 1 MNFAGIEAASAIQG 15

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 47 | 61.0 | 714 | 16 | 09CESS5 |
| 2 | 44 | 57.1 | 225 | 2 | 09RFF6 |
| 3 | 42.5 | 55.2 | 141 | 16 | 09HYR3 |
| 4 | 42 | 54.5 | 369 | 10 | 09Z839 |
| 5 | 42 | 54.5 | 2030 | 10 | 08RZC9 |
| 6 | 41 | 53.2 | 440 | 2 | 09FAD1 |
| 7 | 41 | 53.2 | 443 | 2 | 09AMK7 |
| 8 | 41 | 53.2 | 445 | 16 | 09KV76 |
| 9 | 41 | 53.2 | 340 | 16 | 09CGX6 |
| 10 | 41 | 53.2 | 803 | 5 | 09N978 |
| 11 | 40 | 51.9 | 134 | 8 | 09XMD0 |
| 12 | 40 | 51.9 | 134 | 8 | 09XMD1 |
| 13 | 40 | 51.9 | 323 | 16 | 08UH81 |
| 14 | 40 | 51.9 | 330 | 17 | 027223 |
| 15 | 40 | 51.9 | 379 | 8 | 09B1N0 |
| 16 | 40 | 51.9 | 379 | 8 | 09B1O6 |

| | | | | | | |
|----|----|------|------|----|---------|---------------------|
| 17 | 40 | 51.9 | 379 | 8 | 09B0Y1 | 09b0y1 pteronotus |
| 18 | 40 | 51.9 | 379 | 8 | 09B0U6 | 09b0u6 pteronotus |
| 19 | 40 | 51.9 | 379 | 8 | 09B0S2 | 09b0s2 pteronotus |
| 20 | 40 | 51.9 | 379 | 8 | 09B377 | 09b377 pteronotus |
| 21 | 40 | 51.9 | 379 | 8 | 09B376 | 09b376 pteronotus |
| 22 | 40 | 51.9 | 379 | 8 | 09B375 | 09b375 pteronotus |
| 23 | 40 | 51.9 | 379 | 8 | 09B374 | 09b374 pteronotus |
| 24 | 40 | 51.9 | 379 | 8 | 09B373 | 09b373 pteronotus |
| 25 | 40 | 51.9 | 379 | 8 | 09B372 | 09b372 pteronotus |
| 26 | 40 | 51.9 | 379 | 8 | 09B371 | 09b371 pteronotus |
| 27 | 40 | 51.9 | 379 | 8 | 09B369 | 09b369 pteronotus |
| 28 | 40 | 51.9 | 379 | 8 | 09B368 | 09b368 pteronotus |
| 29 | 40 | 51.9 | 379 | 8 | 09B367 | 09b367 pteronotus |
| 30 | 40 | 51.9 | 379 | 8 | 09B366 | 09b366 pteronotus |
| 31 | 40 | 51.9 | 379 | 8 | 09B365 | 09b365 pteronotus |
| 32 | 40 | 51.9 | 379 | 8 | 09B364 | 09b364 pteronotus |
| 33 | 40 | 51.9 | 379 | 8 | 08WDK6 | 08wdk6 pteronotus |
| 34 | 40 | 51.9 | 419 | 10 | 0941J5 | 0941j5 oryza sativ |
| 35 | 40 | 51.9 | 493 | 16 | 092VB8 | 092vb8 rhizobium m |
| 36 | 40 | 51.9 | 536 | 17 | 09HNR2 | 09hnr2 halobacteri |
| 37 | 40 | 51.9 | 1005 | 12 | 09Q937 | 09q937 shope fibro |
| 38 | 40 | 51.9 | 1006 | 12 | 09Q8R2 | 09q8r2 myxoma viru |
| 39 | 40 | 51.9 | 2535 | 11 | 0912E6 | 0912e6 mus musculu |
| 40 | 40 | 51.9 | 2561 | 11 | 08V1E5 | 08v1e5 mus musculu |
| 41 | 39 | 50.6 | 101 | 17 | 0589E24 | 0589e24 pyrococcus |
| 42 | 39 | 50.6 | 120 | 8 | 08WCZ5 | 08wcz5 arctiocephal |
| 43 | 39 | 50.6 | 120 | 8 | 08WCZ4 | 08wcz4 arctiocephal |
| 44 | 39 | 50.6 | 120 | 8 | 08WCZ3 | 08wcz3 arctiocephal |
| 45 | 39 | 50.6 | 120 | 8 | 08WCZ1 | 08wc21 eumetopias |

ALIGNMENTS

RESULT 1
ID 09CESS5 PRELIMINARY; PRT; 714 AA.
AC 09CESS5;
DT 01-JUN-2001 (TREMBLrel. 17, created)
DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE Glutamine ABC transporter permease and substrate binding protein
DE protein.
GN GMP OR L1759.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RC MEDLINE-21235186; PubMed-11337471;
RA Boletini A., Wincker P., Manger S., Jallion O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403.";
RL Genome Res. 11:731-753(2001).
-1- FUNCTION: PROBABLY PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT
-1- SYSTEM. PROBABLY RESPONSIBLE FOR THE TRANSLLOCATION OF THE
-1- SUBSTRATE ACROSS THE MEMBRANE (BY SIMILARITY).
-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
-1- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-
-1- PROTEIN-DEPENDENT TRANSPORT SYSTEMS.
EMBL: AE006406; AK05857.1; -.
HSSP: P10344; IMDN.
DR InterPro: IPR000515; BPD_transp.
DR InterPro: IPR001311; SBP/glu_receptor.
DR InterPro: IPR001638; SBP_bac_3.
DR Pfam: PF00528; BPD_transp_1.
DR Pfam: PF00497; SBP_bac_3; 2.
DR SMART: SM00062; PBD; 2.
DR PROSITE: PS00402; BPD_TRANS_PNN_MEMBER; 1.
DR PROSITE: PS01039; SBP_BACTERIAL_3; UNKNOWN_1.

KW Transmembrane; Transport; Complete proteome.
 SQ SEQUENCE 714 AA; 78377 MW; 4EBBE5453A968EB CRC64;
 Query Match 61.0%; Score 47; DB 16; Length 714;
 Best Local Similarity 50.0%; Pred. No. 19;
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 WNFAGIEAASAIQ 14
 ||| | :| :| :|
 Db 303 WNFAGIEAASAIQ 316

RESULT 2
 ID 09RFF6 PRELIMINARY; PRT; 225 AA.
 AC 09RFF6: 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DE 01-MAY-2002 (TREMblrel. 20, Last annotation update)
 DE Ured.
 OS Rhodobacter sphaeroides (Rhodospirillum rubrum).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
 OC Rhodobacter.
 OX NCBI_TaxID=1063;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=2.4.1;
 RA Choudhary M., Kaplan S.;
 RT "DNA sequence analysis of the photosynthesis region of Rhodobacter
 sphaeroides 2.4.1.";
 RL Nucleic Acids Res. 0:0-0(2000).
 DR EMBL; AF195122; AAF24251.1; -;
 DR InterPro: IPR002669; Ured.
 DR Pfam; PF01774; Ured; 1.
 SQ SEQUENCE 225 AA; 23417 MW; 59EC44A05802C29F CRC64;

Query Match 57.1%; Score 44; DB 2; Length 225;
 Best Local Similarity 75.0%; Pred. No. 18;
 Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 4 AGIEAASAIQ 15
 ||:||||| |
 Db 178 AGIEAASAIQ 189

RESULT 3
 ID 09HYR3 PRELIMINARY; PRT; 141 AA.
 AC 09HYR3: 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DT 01-OCT-2001 (TREMblrel. 18, Last annotation update)
 DE Hypothetical protein PA3332.
 GN PA3332.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RA MEDLINE=20437337; Pubmed=10984043;
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Kelder J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 opportunistic pathogen.";
 RT Nature 406:959-964(2000).
 DR EMBL; AE004755; AAC06720.1; -;
 DR Hypothetical protein; Complete proteome.

SQ SEQUENCE 141 AA; 15920 MW; 9C0CF669C6C0C5BB CRC64;
 Query Match 55.2%; Score 42.5; DB 16; Length 141;
 Best Local Similarity 40.9%; Pred. No. 20;
 Matches 9; Conservative 2; Mismatches 4; Indels 7; Gaps 1;

OY 1 WNFAGIEAASAIQ 15
 ||| | :| :| :|
 Db 119 WNFAGIEAASAIQ 140

RESULT 4
 ID 09Z839 PRELIMINARY; PRT; 369 AA.
 AC 09Z839: 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
 DE 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE FE-S oxidoreductase.
 GN CPN0513 OR CPJ0513 OR CP0240.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83558;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CWL029;
 RA MEDLINE=99206606; Pubmed=10192388;
 RA Kalman S., Mitchell W., Marathe R., Lamell C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.,
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
 RL Nat. Genet. 21:385-389(1999).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=AR39;
 RA MEDLINE=20150255; Pubmed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Uettermann T., Berry K., Bass S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
 pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RA MEDLINE=20330349; Pubmed=10871362;
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kunara S., Nakazawa T.;
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 from Japan and CWL029 from USA.";
 RL Nucleic Acids Res. 28:2311-2314(2000).
 DR EMBL; AE001636; AAD18653.1; -;
 DR EMBL; AE002184; AAF38105.1; -;
 DR EMBL; AF002546; BAA98719.1; -;
 DR TIGR; CP0240; -;
 DR InterPro: IPR005244; Cons_hypoth423.
 DR TIGR; TIGR00423; TIGR00423; 1.
 KW Complete proteome.

Query Match 54.5%; Score 42; DB 16; Length 369;
 Best Local Similarity 77.8%; Pred. No. 65;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 WNFAGIEAASAIQ 9
 ||| | :| :| :|
 Db 305 WNFAGIEAASAIQ 313

RESULT 5
 ID 08RZC9 PRELIMINARY; PRT; 2030 AA.

```

AC Q8R2C9;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE P0510C12.24 protein.
GN P0510C12.24
OS Oryza sativa (japonica cultivar-group).
OC Euryzota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
OC Spermatophyta: Magnoliophyta: Liliopsida: Poales; Poaceae;
OC Ehrhartoideae: Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0510C12."
RL Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AP003725; BAB90479.1; -.
SQ SEQUENCE 2030 AA; 223434 MW; 26C042AD2D4867ED CRC64;

Query Match 54.5%; Score 42; DB 10; Length 2030;
Best Local Similarity 60.0%; Pred. No. 3.8e+02;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 WNFAGIEAASAI0G 15
DB 400 WNSSGSCNAASAPG 414

RESULT 6
O9FAD1 PRELIMINARY; PRT; 440 AA.
ID O9FAD1
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE CADB.
GN CADB.
OS Moritella japonica.
OC Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;
OC Moritella.
OX NCBI_TaxID=89067;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSK1;
RA Li L., Fujii S., Kato C., Koki H.;
RT "Expression of Lysin Decarboxylase Gene Is Regulated by Pressure in a
RT Deep-Sea Piezophilic Bacterium, Moritella japonica."
RL Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB041735; BAB16717.1; -.
DR InterPro; IPR002293; AA/rel_primeasel.
DR InterPro; IPR004841; Permease.
DR Pfam; PF00324; aa_permeases; 1.
SQ SEQUENCE 440 AA; 45577 MW; DBDEACADFE8958CD9 CRC64;

Query Match 53.2%; Score 41; DB 2; Length 440;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 WNFAGIEAASAI 10
DB 199 WSEFVGESAA 208

RESULT 7
O9AMK7 PRELIMINARY; PRT; 443 AA.
ID O9AMK7
AC O9AMK7
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Cadaverine/Lysine antiporter.

```

```

GN CADB.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RA Rhee J., Lee J., Lee S., Chung S., Rhee J., Choi S., Ryu P.;
RT "Vibrio vulnificus cadda genes."
RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF324470; AAK01663.1; -.
DR InterPro; IPR002293; AA/rel_primeasel.
DR InterPro; IPR004841; Permease.
DR Pfam; PF00324; aa_permeases; 1.
SQ SEQUENCE 443 AA; 46752 MW; B88FD7E18036751 CRC64;

Query Match 53.2%; Score 41; DB 2; Length 443;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 WNFAGIEAASAI 10
DB 199 WSEFVGESAA 208

RESULT 8
O9KV76 PRELIMINARY; PRT; 445 AA.
ID O9KV76
AC O9KV76
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Cadaverine/Lysine antiporter Cadd, putative.
GN VC0280.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Rettell H., Dragoi I., Sellers P.,
RA Ermlaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Uterback T., Fleischmann R.D., Nieman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
DR EMBL; AE004116; AAF93455.1; -.
DR TIGR; VC0280; -.
DR InterPro; IPR002293; AA/rel_primeasel.
DR InterPro; IPR004841; Permease.
DR Pfam; PF00324; aa_permeases; 1.
KV Complete proteome.
SQ SEQUENCE 445 AA; 46907 MW; 3A64F21B9680584C CRC64;

Query Match 53.2%; Score 41; DB 16; Length 445;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 WNFAGIEAASAI 10
DB 199 WSEFVGESAA 208

RESULT 9
O9CGX6 PRELIMINARY; PRT; 540 AA.
ID O9CGX6
AC O9CGX6
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

```

DT 01-MAR-2002 (TReMBLrel. 20, last annotation update)
 DE ABC transporter permease protein.
 GN YJYD OR L10966.
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
 OC Bacteria: Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Streptococcaceae; Lactococcus.
 NC NCB1_TaxID=1360;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21235186; PubMed=11337471;
 RA Bioclin A., Wincker P., Manger S., Jallion O., Malarne K.,
 RT Weissbach J., Ehrlich S.D., Sorokin A.;
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus
 RT lactis ssp. lactis IL1403.";
 RL Genome Res. 11:731-753(2001).
 DR EMBL: AE006331; AK05064.1; -.
 KW Complete proteome.
 SQ SEQUENCE 540 AA; 60283 MW; E896CF783987560E CRC64;

Query Match 53.2%; Score 41; DB 16; Length 540;
 Best Local Similarity 69.2%; Pred. No. 1.4e+02;
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 FAGEAASAIQ 15
 DB 339 FVGNETAASAIRG 351

RESULT 10
 Q9N978 PRELIMINARY; PRT; 803 AA.
 AC Q9N978;
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-JUN-2000 (TReMBLrel. 15, last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, last annotation update)
 DE Possible amino acid transporter.
 GN P1295.10.
 OS Leishmania major.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCB1_TaxID=5664;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FRIEDLIN;
 RA Masny D., Purnelle B., Goffeau A., Ivens A.C., Quail M.,
 RA Rajandream M.A., Barrell B.G.;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FRIEDLIN;
 RX MEDLINE=9816435; PubMed=9477341;
 RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
 RA Smith D.F.;
 RT "A physical map of the Leishmania major Friedlin genome.";
 RL Genome Res. 8:135-145(1998).
 DR EMBL: AL359773; CAB95218.1; -.
 DR InterPro: IPR002293; AA/re1_primeasel.
 DR InterPro: IPR004841; Permease.
 DR Pfam: PF00324; aa_permeases; 1.
 SQ SEQUENCE 803 AA; 84779 MW; 0AEBD76516440558 CRC64;

Query Match 53.2%; Score 41; DB 5; Length 803;
 Best Local Similarity 42.9%; Pred. No. 2.1e+02;
 Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 WNFAGIEAASAIQ 14
 DB 285 WNFANIEAGAVE 298

RESULT 11
 Q9XMD0 PRELIMINARY; PRT; 134 AA.

AC Q9XMD0;
 DT 01-NOV-1999 (TReMBLrel. 12, Created)
 DT 01-NOV-1999 (TReMBLrel. 12, last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, last annotation update)
 DE Cytochrome b (Fragment).
 GN Cytb.
 OS Pteronotus davyi (Davy's naked-backed bat).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Mormoopidae;
 OC Pteronotus.
 OX NCB1_TaxID=94956;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20072938; PubMed=10603267;
 RA Kennedy M., Paterson A.M., Morales J.C., Parsons S., Winnington A.P.,
 RA Spencer H.G.;
 RT "The long and short of it: Branch lengths and the problem of placing
 RT the New Zealand short-tailed bat, Mystacina.";
 RL Mol. Phylogenet. Evol. 13:405-416(1999).
 CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
 CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
 CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
 CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
 CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B565) WHICH ARE NOT COVALENTLY
 CC BOUND TO THE PROTEIN (BY SIMILARITY).
 CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
 CC CYTOCHROME C1 AND THE RISK PROTEIN (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
 DR EMBL: AF144070; AAD34999.1; -.
 DR InterPro: IPR000179; Cyt_b_b6.
 DR Pfam: PF00033; cytochrome_b_n; 1.
 DR PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
 KW Electron transport; Heme; Mitochondrion; Respiratory chain;
 KW Transmembrane.
 FT NON_TER 134
 SQ SEQUENCE 134 AA; 15054 MW; 54FA6146B46D3C03 CRC64;

Query Match 51.9%; Score 40; DB 8; Length 134;
 Best Local Similarity 50.0%; Pred. No. 50;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 WNFAGIEAASAIQ 14
 DB 31 WNFGLAACLAVQ 44

RESULT 12
 Q9XMD1 PRELIMINARY; PRT; 134 AA.
 AC Q9XMD1;
 DT 01-NOV-1999 (TReMBLrel. 12, Created)
 DT 01-NOV-1999 (TReMBLrel. 12, last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, last annotation update)
 DE Cytochrome b (Fragment).
 GN Cytb.
 OS Pteronotus parnellii (Parnell's mustached bat).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Mormoopidae;
 OC Pteronotus.
 OX NCB1_TaxID=59476;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20072938; PubMed=10603267;
 RA Kennedy M., Paterson A.M., Morales J.C., Parsons S., Winnington A.P.,
 RA Spencer H.G.;
 RT "The long and short of it: Branch lengths and the problem of placing
 RT the New Zealand short-tailed bat, Mystacina.";
 RL Mol. Phylogenet. Evol. 13:405-416(1999).
 CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
 CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
 CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL

CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
 CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
 CC BOUND TO THE PROTEIN (BY SIMILARITY).
 CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
 CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
 CC EMBL: AF144071; AAD35000.1; -
 DR InterPro: IPR000179; Cyt_b_b6.
 DR Pfam: PF00033; cytochrome_b_N.1.
 DR PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
 DR Electron transport; Heme; Mitochondrion; Respiratory chain;
 FW Transmembrane.
 KM NON TER 134
 SO SEQUENCE 134 AA; 15086 MW; 5068D4E9AA6D23C4 CRC64;

Query Match 51.9%; Score 40; DB 8; Length 134;
 Best Local Similarity 50.0%; Pred. No. 50;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

1 WNFAGIEAASAIQ 14
 ||| : ||| :
 Db 31 WNFSGSLAACLAQ 44

RESULT 13

ID Q8UH81 PRELIMINARY; PRT; 323 AA.
 AC Q8UH81;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical protein Atu0802.
 GN Atu0802 OR AGR_C1468.
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OC NCBI_TaxID=176299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-21608550; PubMed-11743193;
 RA Wood D.W., Setudal J.C., Kaul R., Monks D.E., Kitaajima J.P.,
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boye D. Sr.,
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
 RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
 RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
 RA Zhang S., Yoo H., Tao H., Biddle P., Jung M., Krespan W., Perry M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Nester E.W.;
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
 RT C58.";
 RL Science 294:2317-2323(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-21608551; PubMed-11743194;
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
 RA Qurollo B., Goldman B.S., Cao Y., Askenazi S., Halling C., Mullin L.,
 RA Houmlet K., Gordon J., Vaudin M., Iatkenouk O., Epp A., Liu F.,
 RA Wollam C., Allinger M., Doughly D., Scott C., Lappas C., Markelz B.,
 RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
 RA Cleto C., Slater S.;
 RT "Genome sequence of the plant pathogen and biotechnology agent
 RT Agrobacterium tumefaciens C58.";
 RL Science 294:2323-2328(2001).
 DR EMBL: AE009047; AAL41818.1; -
 DR EMBL: AE008013; AAK86611.1; -
 KW Hypothetical protein: Complete proteome.
 SO SEQUENCE 323 AA; 35066 MW; C2097F09A00C744E CRC64;

Query Match 51.9%; Score 40; DB 16; Length 323;
 Best Local Similarity 61.5%; Pred. No. 1.2e+02;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 3 FAGIEAASAIQ 15
 ||||| : ||| :
 Db 11 FAGISCAATALAG 23

RESULT 14

ID 027223 PRELIMINARY; PRT; 330 AA.
 AC 027223;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE NA/CA+ EXCHANGING protein related.
 GN MTH155.
 OS Methanobacterium thermoautotrophicum.
 OC Archaeae; Euryarchaeota; Methanobacteria; Methanobacteriales;
 OC Methanobacteriaceae; Methanothermobacter.
 OC NCBI_TaxID=187420;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-DELTA H;
 RX MEDLINE-98037514; PubMed-9371463;
 RA Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., Dubois J.,
 RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert R.,
 RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
 RA Spadatoria R., Vlcare R., Wang Y., Wierzbowski J., Gibson R.,
 RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
 RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
 RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum
 RT deltaH: functional analysis and comparative genomics.";
 RL J. Bacteriol. 179:7135-7155(1997).
 DR EMBL: AE000885; AAB85644.1; -
 DR InterPro: IPR004481; K_NacCaexchg.
 DR InterPro: IPR004837; NaCa_Exmemb.
 DR Pfam: PF01699; Na_Ca_Ex; 2.
 DR TIGRFAMs: TIGR00367; K_NacCaexchang-rel; 1.
 KW Complete proteome.
 SO SEQUENCE 330 AA; 34762 MW; BEE7B029F5A54DF CRC64;

Query Match 51.9%; Score 40; DB 17; Length 330;
 Best Local Similarity 46.7%; Pred. No. 1.3e+02;
 Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

1 WNFAGIEAASAIQ 15
 ||||| : ||| :
 Db 95 WNFAGILGISSVAVG 109

RESULT 15

ID 09BIN0 PRELIMINARY; PRT; 379 AA.
 AC 09BIN0;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DE 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
 DE Cytochrome b.
 GN CYTB.
 OS Pteronotus davyi (Davy's naked-backed bat).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Mormoopidae;
 OC Pteronotus.
 OC NCBI_TaxID=94956;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-27642, AND 35483;
 RA Lewis-Orlit N., Porter C.A., Baker R.J.;
 RT "Molecular Systematics of the Family Mormoopidae (Chiroptera) Based on
 RT Mitochondrial cytochrome b and Nuclear RAG-2 Gene Sequences.";
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE

CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN (BY SIMILARITY).
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
DR EMBL: AF338672; AKK21932.1; -
DR EMBL: AF338670; AKK21930.1; -
DR InterPro: IPR000179; Cyt_b_b6.
DR Pfam: PF000033; cytochrome_b_c1.
DR PROSITE: PS00192; CYTOCHROME_B_HEME_1.
DR PROSITE: PS00193; CYTOCHROME_B_OO; UNKNOWN_1.
KW Electron transport; Heme; Mitochondrion; Respiratory chain;
KW Transmembrane.
SO SEQUENCE 379 AA; 42664 MW; 27FC6661E686E624 CRC64;

Query Match 51.98; Score 40; DB 8; Length 379;
Best Local Similarity 50.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 WNFAGTEAASATQ 14
||| : || -1-
Db 31 WNFGLAACLAAYQ 44

Search completed: July 3, 2003, 14:06:48
Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2003, 14:02:14 ; Search time 6.5 Seconds

(Without alignments)
95.715 Million cell updates/sec

Title: US-09-830-839-6
Sequence: 1 MNFAGIEAASAIAG 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query % | Match Length | ID | Description |
|------------|-------|---------|--------------|----|-------------|
| 1 | 77 | 100.0 | 94 | 1 | ESAB_MYCTU |
| 2 | 43 | 55.8 | 476 | 1 | NOEA_RHIME |
| 3 | 41 | 53.2 | 384 | 1 | PHUB_BACSU |
| 4 | 40 | 51.9 | 294 | 1 | G3P_ESCHL |
| 5 | 40 | 51.9 | 2564 | 1 | SPCO_HUMAN |
| 6 | 39 | 50.6 | 370 | 1 | CYB_EPRCE |
| 7 | 39 | 50.6 | 379 | 1 | CYB_ZALCA |
| 8 | 39 | 50.6 | 425 | 1 | PELD_ERWCH |
| 9 | 39 | 50.6 | 445 | 1 | YUDE_ECOLI |
| 10 | 39 | 50.6 | 990 | 1 | TNP7_ECOLI |
| 11 | 38 | 49.4 | 95 | 1 | ESAB_MYCLE |
| 12 | 38 | 49.4 | 444 | 1 | CADR_ECOLI |
| 13 | 38 | 49.4 | 923 | 1 | YAUD_SCHPO |
| 14 | 38 | 49.4 | 962 | 1 | ARVC_HUMAN |
| 15 | 38 | 49.4 | 969 | 1 | ARVC_MOUSE |
| 16 | 38 | 49.4 | 988 | 1 | TNP2_ECOLI |
| 17 | 37 | 48.1 | 75 | 1 | OCTL_MACEU |
| 18 | 37 | 48.1 | 196 | 1 | RBMS_XENLA |
| 19 | 37 | 48.1 | 335 | 1 | YD59_SCHPO |
| 20 | 37 | 48.1 | 379 | 1 | CYB_CHEMA |
| 21 | 37 | 48.1 | 379 | 1 | CYB_ZAPPR |
| 22 | 37 | 48.1 | 404 | 1 | DGIA_RHILLO |
| 23 | 37 | 48.1 | 475 | 1 | GUINA_CLOCE |
| 24 | 37 | 48.1 | 492 | 1 | HUPR_RHOCA |
| 25 | 37 | 48.1 | 507 | 1 | PSBB_SYNY3 |
| 26 | 37 | 48.1 | 509 | 1 | PSBB_SYNY3 |
| 27 | 37 | 48.1 | 509 | 1 | PSBB_SYNY3 |
| 28 | 37 | 48.1 | 509 | 1 | PSBB_SYNY3 |
| 29 | 36 | 46.8 | 125 | 1 | CYCP_RHOCA |
| 30 | 36 | 46.8 | 145 | 1 | Y008_MYCTU |
| 31 | 36 | 46.8 | 211 | 1 | MT04_SYNY3 |
| 32 | 36 | 46.8 | 211 | 1 | MT04_SYNY3 |
| 33 | 36 | 46.8 | 377 | 1 | FDH_ORYSA |
| | | | | | FDH_HORVU |

ALIGNMENTS

| RESULT 1 | ID | ESAB_MYCTU | STANDARD | PRT | 94 AA. |
|----------|---|-----------------------------------|----------|-----|--------|
| AC | 057165 | 084901 | | | |
| DT | 30-MAY-2000 | (Rel. 39, Created) | | | |
| DT | 30-MAY-2000 | (Rel. 39, Last sequence update) | | | |
| DT | 15-JUN-2002 | (Rel. 41, Last annotation update) | | | |
| DE | 6 kDa early secretory antigenic target (ESAT-6). | | | | |
| GN | ESAT6 OR RV3875 OR MT3989 OR MTW027.10. | | | | |
| OS | Mycobacterium tuberculosis, and | | | | |
| OS | Mycobacterium bovis. | | | | |
| OC | Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae; | | | | |
| OC | Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium. | | | | |
| OX | NCBI_Taxid=1773, 1765; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | SPECIES=M.tuberculosis; STRAIN=H37RV; | | | | |
| RX | MEDLINE=95204931; PubMed=7897219; | | | | |
| RA | Andersen P., Andersen A.B., Sorensen A.L., Nagai S.; | | | | |
| RT | "Recall of long-lived immunity to Mycobacterium tuberculosis infection in mice."; | | | | |
| RL | J. Immunol. 154:3359-3372(1995). | | | | |
| RN | [2] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | SPECIES=M.tuberculosis; STRAIN=H37RV; | | | | |
| RX | MEDLINE=98295987; PubMed=9634230; | | | | |
| RA | Cole S.T., Brosch R., Parthill J., Garnier T., Churcher C., Harris D., | | | | |
| RA | Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekait F., | | | | |
| RA | Badcock R., Basham D., Brown D., Chillingworth T., Connor R., | | | | |
| RA | Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S., | | | | |
| RA | Honrsny T., Jagels K., Kirogh A., Mclean J., Moule S., Murphy L., | | | | |
| RA | Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., | | | | |
| RA | Rutter S., Seeger K., Skelton S., Squares S., Squares R., | | | | |
| RA | Stulsen J.E., Taylor K., Whitehead S., Barrell B.G.; | | | | |
| RT | "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence."; | | | | |
| RL | Nature 393:537-544(1998). | | | | |
| RN | [4] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh; | | | | |
| RA | Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., | | | | |
| RA | Petersen J.F., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E., | | | | |
| RA | Kolonyak J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L., | | | | |
| RA | Bisbal W., Uterback T., Weidman J., Khouri H., Gill J., Mikula A., | | | | |
| RT | "Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains."; | | | | |
| RL | Submitted (APR-2001) to the EMBL/GenBank/DBJ databases. | | | | |

```

RN [5]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis;
RA Singh B., Siddiqui Z., Singh S., Sharma P.;
RT "ESAT-6 gene of a clinical isolate of Mycobacterium tuberculosis from
  India."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis;
RA Mahdires G.G., Sabo P.J., Hickey M.J., Singh D.C., Stover C.K.;
RT "Molecular analysis of genetic differences between Mycobacterium bovis
  BCG and virulent M. bovis."
RL J. Bacteriol. 178:1274-1282(1996).
RN [7]
RP SEQUENCE OF 1-70 FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RA MEDLINE=99061212; PubMed=9846755;
RA Berthet F.-X., Rasmussen P.B., Rosenkrands I., Andersen P.,
  Gicquel B.;
RT "A Mycobacterium tuberculosis operon encoding ESAT-6 and a novel
  low-molecular-mass culture filtrate protein (CFP-10).";
RL Microbiology 144:3195-3203(1998).
CC -1- FUNCTION: NOT KNOWN. ELICITS HIGH LEVEL OF INF-GAMMA FROM MEMORY
  EFFECTOR CELLS DURING THE FIRST PHASE OF A PROTECTIVE IMMUNE
  RESPONSE.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE ESAT6 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U34848; AAC44033.1; -
DR EMBL: X79562; CAA56099.1; -
DR EMBL: AL022120; CAA17967.1; -
DR EMBL: AF420491; AAL16896.1; -
DR EMBL: AE007190; AAK48357.1; -
DR EMBL: AF004671; AAC83446.1; -
DR TIGR: MT3989; -
DR Tuberculist: RV3875; -
KW Antigen: Complete.proteome.
FT INIT_MET 0
SQ SEQUENCE 94 AA; 9773 MW; 19245B0EC478BC84 CRC64;

Query Match 100.0%; Score 77; DB 1; Length 94;
Best Local Similarity 100.0%; Pred. No. 5.1e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WNFAGIEAASATOG 15
Db 5 WNFAGIEAASATOG 19

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RCR2011 / SU47;
RX MEDLINE=96111489; PubMed=8801423;
RA Ardourel M., Lortet G., Maillet F., Roche P., Truchet G.,
  Prome J.-C., Rosenberg C.;
RT "In Rhizobium meliloti, the operon associated with the nod box n5
  comprises nodJ, noeA and noeB, three host-range genes specifically
  required for the nodulation of particular Medicago species.";
RL Mol. Microbiol. 17:687-699(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
  Barloy-Hubler F., Bowser L., Capela D., Gilbert F., Gouzy J.,
  Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
  Kaiman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
  Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
  Sinorhizobium meliloti pSymA megaplasmid.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
CC -1- FUNCTION: NOT KNOWN. DOES NOT SEEM TO PARTICIPATE IN NOD FACTOR
  SYNTHESIS BUT REQUIRED FOR NODULATION ON SOME SPECIFIC MEDICAGO
  SPECIES SUCH AS M.LITTORALIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U26430; AAC44091.1; -
DR EMBL: AE007232; AAK65074.1; -
DR InterPro: IPR000051; SAM_bind.
KW Plasmid; Nodulation; Complete proteome.
SQ SEQUENCE 476 AA; 53673 MW; EF0A4009B5F4965E CRC64;

Query Match 55.8%; Score 43; DB 1; Length 476;
Best Local Similarity 70.0%; Pred. No. 9.6;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 WNFAGIEAANA 10
Db 101 WNFAGIEAANA 110

RESULT 3
FHUB_BACSU
ID FHUB_BACSU STANDARD; PRT; 384 AA.
AC P49936;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ferrichrome transport system permease protein fhub.
GN FHUB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Schneider R., Hantke K.;
RT Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98304083; PubMed=9639930;
RA Wipat A., Brignell C.S., Guy J.B., Rose M., Emerson P.T.,
  Harwood C.R.;
RT "The ysa-yvga (293 degrees - 289 degrees) region of the Bacillus

```

RT subtilis chromosome containing genes involved in metal ion uptake and
 RT a putative sigma factor.";
 RL Microbiology 144:1593-1600(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE-98044033; PubMed-9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bortner M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codan J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
 RA Medina N., Meliade R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogilwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Seriguchi J., Sekowska A., Serot S.J., Serro P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Tarpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassart A.,
 RA Viari A., Wambuit R., Wedler H., Wedler H., Welteneeger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis.";
 RL Nature 390:249-256(1997).
 RN [4]
 RP SEQUENCE OF 1-38 FROM N.A.
 RC STRAIN-168 / Marburg;
 RX MEDLINE-93268086; PubMed-838528;
 RA Schneider R., Hantke K.;
 RT "Iron-hydroxamate uptake systems in Bacillus subtilis: identification
 RT of a lipoprotein as part of a binding protein-dependent transport
 RT system.";
 RL Mol. Microbiol. 8:111-121(1993).
 CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS INVOLVED IN THE HIGH-
 CC AFFINITY TRANSPORT OF IRON(III)-FERRICHYDROXY INTO THE CELL.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO THE BINDING-PROTEIN-DEPENDENT TRANSPORT
 CC SYSTEM PERMEASE FAMILY. FECCD SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X93092; CA63643.1; -;
 CC EMBL; AJ223978; CA611720.1; -;
 CC EMBL; Z99121; CAB15337.1; -;
 CC EMBL; M87283; -; NOT_ANNOTATED_CDS.
 CC Subtilisin; Bg11389; fnub.
 CC InterPro: IPR000515; BPD_TRANSP.
 CC InterPro: IPR000522; FECCD.
 CC Pfam: PF01032; FECCD. 1.
 CC ProDom: PD001557; FECCD. 1.
 CC PROSITE; PS00402; BPD_TRANSP_INN_MEMBER; FALSE_NEG.
 CC Transprot; Iron transport; Transmembrane; Complete proteome.
 KW TRANSMEM 58 78 POTENTIAL.

FT TRANSMEM 115 135 POTENTIAL.
 FT TRANSMEM 154 174 POTENTIAL.
 FT TRANSMEM 176 196 POTENTIAL.
 FT TRANSMEM 202 222 POTENTIAL.
 FT TRANSMEM 243 263 POTENTIAL.
 FT TRANSMEM 296 316 POTENTIAL.
 FT TRANSMEM 330 350 POTENTIAL.
 FT TRANSMEM 357 377 POTENTIAL.
 SQ SEQUENCE 384 AA; 40720 MW; 7E4D75CAB36AB8FB CRC64;
 Query Match 53.2%; Score 41; DB 1; Length 384;
 Best Local Similarity 46.7%; Pred. No. 17;
 Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
 QY 1 WNFAGIEAASAIQ 15
 Db 174 WSFAGAGASTVMG 188
 RESULT 4
 G3P_ESCBL STANDARD; PRT; 294 AA.
 ID G3P_ESCBL
 AC P24749;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.12) (GAPDH)
 DE (Fragment).
 DE
 GN GAP.
 OS Escherichia blattae.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=563;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 29907, ATCC 33429, and ATCC 33430;
 RX MEDLINE-92065252; PubMed-1955870;
 RA Lawrence J.G., Ochman H., Harcl D.L.;
 RT "Molecular and evolutionary relationships among enteric bacteria.";
 RL J. Gen. Microbiol. 137:1911-1921(1991).
 CC -1- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate + phosphate +
 CC NAD(+) = 3-phospho-D-glyceroyl phosphate + NADH.
 CC -1- PATHWAY: Second phase of glycolysis; first step.
 CC -1- SUBUNIT: HOMOTETRAMER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE
 CC DEHYDROGENASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M63358; AAA23837.1; -;
 CC EMBL; M63359; AAA23852.1; -;
 CC EMBL; M63360; AAA23855.1; -;
 CC HSP; P06977; IGAD.
 CC InterPro: IPR000173; GAP_dhdrogenase.
 CC Pfam: PF00044; gpdh. 1.
 CC ProSITE; PS00711; GAPDH. 1.
 CC Glycolysis; Oxidoreductase; NAD.
 KW NON_TER 1
 FT BINDING 135 135 GLYCERALDEHYDE 3-PHOSPHATE.
 FT ACT_SITE 162 162 ACTIVATES THIOL GROUP DURING CATALYSIS.
 FT NON_TER 294 294
 SQ SEQUENCE 294 AA; 31372 MW; 1B562CB38177DB8 CRC64;
 Query Match 51.9%; Score 40; DB 1; Length 294;
 Best Local Similarity 58.3%; Pred. No. 19;

Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MNFAGIEAASA 12
| | | | |
Db 70 MNFAGIEVVAEA 81

RESULT 5
SPEC. HUMAN STANDARD; PRT: 2564 AA.
ID SPCO_HUMAN Q9H254: Q9H254: Q9H1K7; Q9H1K8; Q9H1K9;
AC Q9H254: Q9H254: Q9H1K7; Q9H1K8; Q9H1K9;
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Spectrin beta chain, brain 3 (Spectrin, non-erythroid beta chain 3)
DE (Beta-IV spectrin).
GN SPTBN4 OR SPTBN3 OR KIAA1642.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=21316449; PubMed=11294830;
RA Tse W.T., Tang J., Jin O., Korsgren C., John K.M., Kung A.L.,
RA Gwynn B., Peters L.L., Lux S.E.;
RT "A new spectrin, beta-IV, has a major truncated isoform that
RT associates with promyelocytic leukemia protein nuclear bodies and the
RT nuclear matrix."
RL J. Biol. Chem. 276:23974-23985(2001).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 TO 4).
RX MEDLINE=20539976; PubMed=11086001;
RA Bergs S., Aguijaro D., Dirix R.J., Maksimova E., Stabach P.,
RA Hermet J.-M., Zhang J.-P., Philbrick W., Slepnev V., Ort T.,
RA Solimene M.;
RT "BetaIV spectrin, a new spectrin localized at axon initial segments
RT and nodes of ranvier in the central and peripheral nervous system."
RL J. Cell Biol. 151:985-1002(2000).
RN [3]
RP SEQUENCE OF 386-2382 FROM N.A. (ISOFORM 1).
RX TISSUE=Brain;
RA MEDLINE=20450683; PubMed=10997877;
RA Nagase T., Kikuno R., Nakayama M., Hirasawa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.
RT XVII. The complete sequences of 100 new cDNA clones from brain which
RT code for large proteins in vitro."
RL DNA Res. 7:273-281(2000).
CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1 (shown here), 2, 3 and 4; are
CC produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: Abundantly expressed in brain and pancreatic
CC islets.
CC -1- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.
CC -1- SIMILARITY: CONTAINS 2 CALPONTIN-HOMOLOGY (CH) DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: CONTAINS 18 SPECTRIN REPEATS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF311855; AAG42473.1; -
DR EMBL: AF082075; AAG38874.1; -
DR EMBL: AF004226; AAF93171.1; -
DR EMBL: AY004226; AAF93172.1; -
DR EMBL: AY004227; AAF93173.1; -
DR EMBL: AB046862; BAB13468.1; -
DR HSSP: 001082; 1BKR.
DR Gene: HGNC:14896; SPTBN4.

DR MIM: 606214; -
DR InterPro: IPR001589; Actbind_actin.
DR InterPro: IPR001715; Calponin-like.
DR InterPro: IPR001849; PH.
DR InterPro: IPR002017; Spectrin.
DR InterPro: IPR001605; Spectrin_PH.
DR Pfam: PF00307; CH; 2.
DR Pfam: PF00435; spectrin; 56.
DR PRINTS: PR00683; SPECTRINPH.
DR SMART: SM00233; CH; 2.
DR SMART: SM00233; PH; 1.
DR PROSITE: PS00019; ACTININ_1; 1.
DR PROSITE: PS00020; ACTININ_2; 1.
DR PROSITE: PS50021; CH; 2.
DR PROSITE: PS50003; PH_DOMAIN; 1.
KW Cytoskeleton; Membrane; Repeat; Actin-binding; Capping protein;
KW Alternative splicing.
FT DOMAIN 1 282
FT 1 165
FT DOMAIN 180 282
FT REPEAT 309 354
FT REPEAT 398 419
FT REPEAT 429 533
FT REPEAT 535 642
FT REPEAT 644 771
FT REPEAT 773 879
FT REPEAT 881 985
FT REPEAT 1019 1086
FT REPEAT 1088 1197
FT REPEAT 1199 1303
FT REPEAT 1305 1408
FT REPEAT 1410 1513
FT REPEAT 1515 1619
FT REPEAT 1621 1725
FT REPEAT 1727 1832
FT REPEAT 1834 1940
FT REPEAT 1942 2046
FT REPEAT 2048 2107
FT DOMAIN 2418 2527
FT VARSPLIC 1 1257
FT VARSPLIC 1258 1286
FT VARSPLIC 1287 1309
FT VARSPLIC 1310 2564
FT VARSPLIC 2113 2154
FT VARSPLIC 2155 2564
FT CONFLICT 604 608
FT CONFLICT 714 714
FT CONFLICT 1189 1189
FT CONFLICT 1193 1193
FT CONFLICT 1331 1331
SQ SEQUENCE 2564 AA; 288982 MW; 52CDE7D11D601ECC CRC64;
Query Match 51.9%; Score 40; DB 1; Length 2564;
Best local Similarity 42.9%; Pred. No. 14e+02;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 1 MNFAGIEAASA 14
| | | | |
1022 WRLSGEALQALE 1035

RESULT 6
CYB_EPICE STANDARD; PRT: 370 AA.
ID CYB_EPICE
AC Q48043;
DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cytochrome B.
 GN MTCYB OR COB OR CYTB.
 OS Epilarches cenchria.
 OC Mitochondrion.
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Lepidodonta; Squamata; Scleroglossa; Serpentes; Henophidia; Boidae;
 CC Boidae; Epicrates.
 OX NCBI_TaxID=51743;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Campbell B.N.;
 RT "Hic Surt Serpentes - molecular phylogenetics and the Boidae
 (Serpentes: Boidae).";
 RL Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
 CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
 CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
 CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
 CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
 CC BOUND TO THE PROTEIN (BY SIMILARITY).
 CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
 CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U69777; AAC01811.1; -
 DR EMBL: U69778; AAC01812.1; -
 DR InterPro: IPR000179; CYL_b_b6.
 DR Pfam: PF000032; cytochrome_b_c1.
 DR Pfam: PF000033; cytochrome_b_n; 1.
 DR PROSITE: PS00193; CYTOCHROME_B_QQ; 1.
 DR PROSITE: PS00192; CYTOCHROME_B_HEME; FALSE NEG.
 DR Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
 KM Heme.
 FT METAL 75 75 IRON 1 (HEME B562 AXIAL LIGAND).
 FT METAL 89 89 IRON 2 (HEME B566 AXIAL LIGAND).
 FT METAL 174 174 IRON 2 (HEME B562 AXIAL LIGAND).
 FT METAL 188 188 IRON 1 (HEME B566 AXIAL LIGAND).
 SQ SEQUENCE 370 AA; 41744 MW; A42257B25EDDE2B CRC64;
 Query Match 50.6%; Score 39; DB 1; Length 370;
 Best Local Similarity 50.0%; Pred. No. 35;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 OY 1 WNFAGIEAASAIQ 14
 DB 23 WNFSGMLLACSLAQ 36
 RESULT 7
 CYB_ZALCA STANDARD; PRT; 379 AA.
 AC Q36266;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Cytochrome B
 GN MTCYB OR COB OR CYTB.
 OS Zalophus californianus (California sealion).
 OC Mitochondrion.
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Carnivora; Planipecta; Otariidae; Zalophus.
 OX NCBI_TaxID=9704;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE=95230701; PubMed=7714914;
 RA Arason U., Bodin K., Gulberg A., Ledje C., Mouchaty S.;
 RT "A molecular view of pliniped relationships with particular emphasis
 on the true seals.";
 RL J. Mol. Evol. 40:78-85(1995).
 CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
 CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
 CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
 CC COUPLED TO ATP SYNTHESIS.
 CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
 CC BOUND TO THE PROTEIN.
 CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
 CC CYTOCHROME C1 AND THE RIESKE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X82310; CAA57753.1; -
 DR HSP: P00157; 1BE3.
 DR InterPro: IPR000179; CYL_b_b6.
 DR Pfam: PF000032; cytochrome_b_c1.
 DR Pfam: PF000033; cytochrome_b_n; 1.
 DR PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
 DR PROSITE: PS00193; CYTOCHROME_B_QQ; 1.
 DR Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
 KM Heme.
 FT METAL 83 83 IRON 1 (HEME B562 AXIAL LIGAND).
 FT METAL 97 97 IRON 2 (HEME B566 AXIAL LIGAND).
 FT METAL 182 182 IRON 2 (HEME B562 AXIAL LIGAND).
 FT METAL 196 196 IRON 1 (HEME B566 AXIAL LIGAND).
 SQ SEQUENCE 379 AA; 42520 MW; F74CD03BA27C24D4 CRC64;
 Query Match 50.6%; Score 39; DB 1; Length 379;
 Best Local Similarity 50.0%; Pred. No. 36;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 OY 1 WNFAGIEAASAIQ 14
 DB 31 WNFSGMLLACSLAQ 44
 RESULT 8
 PELL_ERWCH STANDARD; PRT; 425 AA.
 ID PELL_ERWCH
 AC Q47473; O59421;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Pectate lyase I precursor (EC 4.2.2.2) (Pectate transeliminase).
 GN PELL.
 OS Erwinia chrysanthemi.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Pectobacterium.
 OX NCBI_TaxID=556;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EC16;
 RX MEDLINE=95362691; PubMed=7635842;
 RA Alfano J.R., Ham J.H., Collmer A.;
 RT "Use of Tnshact to clone a pel gene encoding a highly alkaline,
 RT asparagine-rich pectate lyase isozyme from an Erwinia chrysanthemi
 RT EC16 mutant with deletions affecting the major pectate lyase
 RT isozymes";
 RL J. Bacteriol. 177:4553-4556(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3937;

RX MEDLINE-96020649; PubMed-8577252;
 RA Lojkowski E., Masclaux C., Boccarda M., Robert-Baudouy J.,
 RT "Characterization of the *pel* gene encoding a novel pectate lyase of
 RL *Microbium* 16.1183-1195(1995).
 CC
 CC -1- FUNCTION: PRESENTS AN ENDO-CLEAVING ACTIVITY ON POLYGALACTURONATE
 CC OR PARTIALLY METHYLATED PECTIN.
 CC -1- CATALYTIC ACTIVITY: Eliminative cleavage of pectate to give
 CC oligosaccharides with 4-deoxy-alpha-D-gluc-4-enuronosyl groups at
 CC their non-reducing ends.
 CC -1- COFACTOR: REQUIRES CALCIUM.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- INDUCTION: BY PECTIC CATABOLIC PRODUCTS AND AFFECTED BY GROWTH
 CC PHASE, TEMPERATURE, IRON STARVATION, OSMOLARITY, ANAEROBIOSIS,
 CC NITROGEN STARVATION AND CATABOLITE REPRESSION.
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 9.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: I42248; AAA99476.1; -;
 DR EMBL: X81136; CAA57041.1; -;
 KM Lyase; Signal; Calcium.
 FT STGNAL 1 25
 FT CHAIN 26 425
 FT VARIANT 49 425
 FT VARIANT 49 425
 FT VARIANT 55 55
 FT VARIANT 59 59
 FT VARIANT 64 64
 FT VARIANT 70 70
 FT VARIANT 103 105
 FT VARIANT 147 147
 FT VARIANT 160 160
 FT VARIANT 219 219
 FT VARIANT 224 224
 FT VARIANT 261 261
 FT VARIANT 264 264
 FT VARIANT 328 328
 FT VARIANT 343 348
 FT VARIANT 362 362
 FT VARIANT 378 378
 FT VARIANT 390 390
 FT VARIANT 396 396
 FT VARIANT 425 425
 SQ SEQUENCE 425 AA; 45537 MW; BE2557332CF91F7 CRC64;
 Query Match 50.6%; Score 39; DB 1; Length 425;
 Best Local Similarity 46.7%; Pred. No. 40;
 Matches 7; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OX NCBI_TaxID=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE-95334362; PubMed-7610040;
 RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
 RT Blattner F.R.;
 RT "Analysis of the *Escherichia coli* genome VI: DNA sequence of the
 RL region from 92.8 through 100 minutes."
 RN Nucleic Acids Res. 23:2105-2119(1995).
 RN [2]
 RP REVISIONS.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE-97426317; PubMed-9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12."
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE-21074935; PubMed-11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., DiMantola E.T., Potamosis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7."
 RL Nature 409:529-533(2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / RIMD 0509952;
 RX MEDLINE-21156231; PubMed-11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kumara S., Shida T., Hattori M., Shingawa H.;
 RT "Complete genome sequence of enterohemorrhagic *Escherichia coli*
 RT O157:H7 and genomic comparison with a laboratory strain K-12."
 RL DNA Res. 8:11-22(2001).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (Potential).
 CC -1- SIMILARITY: BELONGS TO THE ARCD/CAD/GERAB/LYSI FAMILY OF
 CC PERMEASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U14003; AAA97015.1; ALT_FRAME.
 DR EMBL: U14003; AAA97014.1; ALT_FRAME.
 DR EMBL: AE000484; AAC77076.1; -;
 DR EMBL: AE0005644; AAC59314.1; -;
 DR EMBL: AF002568; BAB38520.1; -;
 DR EcGene; EG12462; yjyde.
 DR InterPro; IPR002293; AA/re1_pmease1.
 DR Pfam; PF00324; aa_permeases; 1.
 KM Hypothetical protein; Transport; Transmembrane; Inner membrane;
 FT Complete proteome. 30
 FT TRANSMEM 10
 FT TRANSMEM 39 59
 FT TRANSMEM 99 119
 FT TRANSMEM 123 143
 FT TRANSMEM 153 173
 FT TRANSMEM 197 217
 FT POTENTIAL.
 FT POTENTIAL.
 FT POTENTIAL.
 FT POTENTIAL.

FT TRANSMEM 226 246 POTENTIAL.
 FT TRANSMEM 276 296 POTENTIAL.
 FT TRANSMEM 322 342 POTENTIAL.
 FT TRANSMEM 356 376 POTENTIAL.
 FT TRANSMEM 386 406 POTENTIAL.
 FT TRANSMEM 409 429 POTENTIAL.
 SQ SEQUENCE 445 AA; 46842 MW; 359F70C489A20663 CRC64;

Query Match 50.6%; Score 39; DB 1; Length 445;
 Best Local Similarity 50.0%; Pred. No. 42;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 1 MNFAGIEAASA 12
 Db 202 MSFIEGASASVA 213

RESULT 10
 ID TNP7_ECOLI STANDARD; PRT; 990 AA.
 AC P13694;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transposase for transposon Tn3926.
 GN TNP4.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=89160328; PubMed=2537961;
 RA Turner A.K., Grinstead J.;
 RT "DNA sequence of the transposase gene of the class II transposon, Tn3926";
 RL Nucleic Acids Res. 17:1757-1757(1989).
 CC -1- FUNCTION: REQUIRED FOR TRANSPOSITION OF TRANSPOSON TN3926.
 CC -1- SIMILARITY: BELONGS TO THE TRANSPOSASE FAMILY 7.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X14236; CAA32453.1; -
 DR PIR: S03285; TOEC26.
 DR InterPro: IPR002513; Transposase_7.
 DR Pfam: PF01526; Transposase_7.1.
 KW Transposable element; transposition; DNA-binding; DNA recombination.
 SQ SEQUENCE 990 AA; 111073 MW; 0664D1EFED038F606 CRC64;

Query Match 50.6%; Score 39; DB 1; Length 990;
 Best Local Similarity 46.7%; Pred. No. 86;
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 MNFAGIEAASA 15
 Db 927 MNTVLERATSALRG 941

RESULT 11
 ID ESAG_MYCLE STANDARD; PRT; 95 AA.
 AC O50206; O33083;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 6 kda early secretory antigenic target homolog (ESAT-6-like protein)

DE (L-ESAT).
 GN ESAT6 OR ESX OR I45 OR ML0049 OR MLCB628.12C.
 OS Mycobacterium leprae.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Wields B., Notenhoom T., Naafs B., Offringa R., Ottenhoff T.;
 RL Submitted (Aug-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN-TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eigemeier K., Parthill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Fellwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajadream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Bartell B.G.;
 RT "Massive gene decay in the leprosy bacillus";
 RL Nature 409:1007-1011(2001).
 CC -1- SUBCELLULAR LOCATION: Secreted (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE ESAT6 FAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: X90946; CAA62441.1; -
 DR EMBL: Y14967; CAA75200.1; -
 DR EMBL: AL583917; CAC29557.1; -
 DR Leptoma; ML0049; -
 KW Complete proteome.
 FT CONFLICT 55 R -> Q (IN REF. 1).
 FT CONFLICT 90 M -> T (IN REF. 1).
 SQ SEQUENCE 95 AA; 10465 MW; B1526F78CB2B8A1 CRC64;

Query Match 49.4%; Score 38; DB 1; Length 95;
 Best Local Similarity 33.3%; Pred. No. 15;
 Matches 5; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY 1 MNFAGIEAASA 15
 Db 5 WHFPALGAVNEI 19

RESULT 12
 ID CADR_ECOLI STANDARD; PRT; 444 AA.
 AC P23891;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable cadaverine/lysine antiporter.
 GN CADR OR B4132 OR Z5735 OR ECS5114.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92105022; PubMed=1370290;
 RA Watson N., Dunyak D.S., Rosey E.L., Slonczewski J.L., Olson E.R.;
 RT "Identification of elements involved in transcriptional regulation of

the Escherichia coli cad operon by external pH.";
 J. Bacteriol. 174:530-540(1992).
 [2]
 SEQUENCE FROM N.A.
 MEDLINE-92210511; PubMed-1556085;
 Meng S.-Y., Bennett G.N.;
 "Nucleotide sequence of the Escherichia coli cad operon: a system for
 neutralization of low extracellular pH.";
 J. Bacteriol. 174:2659-2669(1992).
 [3]
 SEQUENCE FROM N.A.
 STRAIN-K12 / MG1655;
 MEDLINE-95334362; PubMed-7610040;
 Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
 Blattner F.R.;
 "Analysis of the Escherichia coli genome VI: DNA sequence of the
 region from 92.8 through 100 minutes.";
 Nucleic Acids Res. 23:2105-2119(1995).
 [4]
 SEQUENCE FROM N.A.
 STRAIN-0157:H7 / EDL933 / ATCC 700927;
 MEDLINE-21074935; PubMed-11206551;
 Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
 Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 Welch R.A., Blattner F.R.;
 "Genome sequence of enterohemorrhagic Escherichia coli O157:H7.";
 Nature 409:529-533(2001).
 [5]
 SEQUENCE FROM N.A.
 STRAIN-0157:H7 / RIMD 0509952;
 MEDLINE-21156231; PubMed-11258796;
 Havesht T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 Ikeda T., Takami H., Honda T., Sasakiwa C., Ogasawara N., Yasunaga T.,
 Kuhnra S., Shiba T., Hattori M., Shingawa H.;
 "Complete genome sequence of enterohemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12.";
 DNA Res. 8:11-22(2001).
 [6]
 FUNCTION: PROBABLE CADHERINE/LYSINE ANTIPODTER OR PART OF IT.
 CELLULAR LOCATION: Integral membrane protein. Inner membrane.
 INDUCTION: AT LOW PH IN THE PRESENCE OF LYSINE AND ACETIVES
 MAXIMAL LEVEL UNDER ANAEROBIC CONDITIONS.
 SIMILARITY: BELONGS TO THE ARCD/CADH/GERAB/LYSI FAMILY OF
 PERASES.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL: M67452; AAA23532.1; -
 EMBL: M76411; AAA23535.1; -
 EMBL: U14003; AAA97032.1; -
 EMBL: AE000486; AAC77093.1; -
 EMBL: AE005647; AAG59332.1; -
 EMBL: AP002568; BAB38537.1; -
 PIR: B41968; B41968.
 PIR: A41842; A41842.
 Ecogene: EG10132; cadB.
 InterPro: IPR002293; AA/rel_pmeasel.
 InterPro: IPR004841; Permease.
 InterPro: IPR004754; R/O_antiport.
 Pfam: PF00324; aa-permeases.1.
 TIGRFAMs: TIGR00905; 2A0302.1.
 Transport: Antiport; Amino-acid transport; Transmembrane;
 Inner membrane; Complete proteome.
 TRANSMEM 7 27 POTENTIAL.

FT TRANSMEM 35 55 POTENTIAL.
 FT TRANSMEM 68 88 POTENTIAL.
 FT TRANSMEM 95 115 POTENTIAL.
 FT TRANSMEM 123 143 POTENTIAL.
 FT TRANSMEM 149 169 POTENTIAL.
 FT TRANSMEM 193 213 POTENTIAL.
 FT TRANSMEM 222 242 POTENTIAL.
 FT TRANSMEM 250 270 POTENTIAL.
 FT TRANSMEM 273 293 POTENTIAL.
 FT TRANSMEM 323 343 POTENTIAL.
 FT TRANSMEM 354 374 POTENTIAL.
 FT TRANSMEM 384 404 POTENTIAL.
 FT TRANSMEM 405 425 POTENTIAL.
 SQ SEQUENCE 444 AA; 46665 MW; E87913B449B0500A CRC64;
 Query Match 49.4%; Score 38; DB 1; Length 444;
 Best Local Similarity 60.0%; Pred. No. 61;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 OY 1 WNFAGIEAAA 10
 Db 198 WAFVGESEA 207
 ID YAVU_SCHPO STANDARD; PRT; 923 AA.
 YAVU_SCHPO
 AC 01-0CT-1996 (rel. 34, Created)
 DT 01-0CT-1996 (rel. 34, Last sequence update)
 DT 15-JUN-2002 (rel. 41, Last annotation update)
 DE Hypothetical protein C26A3.10 in chromosome I.
 GN SPAC26A3.10.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972;
 RX MEDLINE-21848401; PubMed-11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkhardt G., Aert R., Robben J., Gymnopoulos B.,
 RA Wellens J., Vansteirels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Mambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leclerc V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garçon A., Rhode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Bento J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe.";
 RL Nature 415:871-880(2002).
 CC FUNCTION: GTPase-activating protein for the ADP ribosylation
 factor family (potentially).
 CC SIMILARITY: CONTAINS 1 ARP-GAP DOMAIN.
 CC SIMILARITY: CONTAINS 1 PH DOMAIN.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: 269240; CAA93233.1; -
 CC HSSP: 062261; IPRN.
 CC InterPro: IPR001849; PH.
 CC InterPro: IPR001164; hrp1-like.
 CC Pfam: PF00169; PH; 1.
 CC SMART: SM001412; ArfGAP; 1.
 CC SMART: SM00105; ArfGAP; 1.
 CC SMART: SM00233; PH; 1.
 CC PROSITE: PS50115; ARFGAP; 1.
 CC PROSITE: PS50003; PH_DOMAIN; 1.
 CC Hypothetical protein; GTPase activation; Zinc-finger.
 CC DOMAIN 526 632 PH.
 CC DOMAIN 714 837 ARF-GAP.
 CC ZN_FING 730 754 C4-TYPE.
 CC SEQUENCE 923 AA; 104496 MW; AB05E0F369E0AAB3 CRC64;
 SQ
 Query Match 49.4%; Score 38; DB 1; Length 923;
 Best Local Similarity: 42.9%; Pred. No. 1.2e+02;
 Matches 6; Conservative 6; Mismatches 2; Indels 0; Gaps 0;
 QY 2 NFAGTEAASATOG 15
 DB 831 DFAGVADATLLEG 844
 ID ARVC_HUMAN STANDARD; PRT; 962 AA.
 AC 000192;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Armadillo repeat protein deleted in velo-cardio-facial syndrome.
 GN ARVCF.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (SHORT AND LONG ISOFORMS).
 RX MEDLINE=97271559; PubMed=9126485;
 RA Strockin H., O'Donnell H., Dasgupta R., Halford S., St Jore B.,
 RA Puech A., Parimoo S., Morrow B., Skoultsch A., Weissman S.,
 RA Scambler P., Kucherlapati R.;
 RT "Identification of a new human catenin gene family member (ARVCF) from
 RT the region deleted in velo-cardio-facial syndrome.";
 RL Genomics 41:75-83(1997).
 CC -1- FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS AT ADHERENS
 CC JUNCTIONS.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A SHORT FORM AND A LONG FORM
 CC (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: FOUND IN ALL THE EXAMINED TISSUES INCLUDING
 CC HEART BRAIN, LIVER AND KIDNEY. FOUND AT LOW LEVEL IN LUNG.
 CC -1- DISEASE: GENE DELETED IN VELO-CARDIO-FACIAL SYNDROME (VCFS) AND IS
 CC HEMIZYGOUS IN ALL VCFS PATIENTS WITH INTERSTITIAL DELETIONS. THIS
 CC HEMIZYGOUSITY MAY PLAY A ROLE IN THE ETIOLOGY OF SOME OF THE
 CC PHENOTYPES ASSOCIATED WITH VCFS CHARACTERIZED BY A WIDE SPECTRUM
 CC PHENOTYPES, INCLUDING CONOTRUNCAL HEART DEFECTS, CLEFT PALATE AND
 CC FACIAL DYSMORPHOLOGY.
 CC -1- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 10 ARM REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: 051269; AAC51202.1; -
 CC HSSP: P35222; IG3f.
 CC Genew: HGNC:728; ARVCF.
 CC MIM: 602269; -
 CC MIM: 192430; -
 CC InterPro: IPR000225; Armadillo.
 CC Pfam: PF00514; Armadillo_seg; 6.
 CC SMART: SM00185; ARM; 4.
 CC PROSITE: PS50176; ARM_REPEAT; 3.
 CC Cell adhesion: Cytoskeleton; structural protein; Repeat; Coiled coil;
 CC Alternative splicing.
 CC DOMAIN 8 46
 CC DOMAIN 607 623 COILED COIL (POTENTIAL).
 CC DOMAIN 348 387 POLY-ARG.
 CC REPEAT 348 387 ARM 1.
 CC REPEAT 390 429 ARM 2.
 CC REPEAT 433 467 ARM 3.
 CC REPEAT 468 508 ARM 4.
 CC REPEAT 526 565 ARM 5.
 CC REPEAT 575 622 ARM 6.
 CC REPEAT 646 686 ARM 7.
 CC REPEAT 699 738 ARM 8.
 CC REPEAT 739 781 ARM 9.
 CC REPEAT 782 826 ARM 10.
 CC VARSPLIC 1 69 MEDICVNSASIIASVKEQAEFRERLTRADEERRHVALQL
 CC ERAOQPGWVSGMGSGGQPLPMAMQQLVL -> MPAEELR
 CC (IN SHORT ISOFORM).
 SQ SEQUENCE 962 AA; 104641 MW; 7A41814A022FF2B1 CRC64;
 ID ARVC_MOUSE STANDARD; PRT; 969 AA.
 AC P98203;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Armadillo repeat protein deleted in velo-cardio-facial syndrome
 DE homolog (Fragment).
 GN ARVCF.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20512094; PubMed=11058098;
 RA Kaufmann U., Zupplinger C., Waidler Z., Rudiger M., Urbich C.,
 RA Martin B., Jockusch B.M., Eppenberger H., Starzinski-Powitz A.;
 RT "The armadillo repeat region targets ARVCF to cadherin-based cellular
 RT junctions.";
 RL J. Cell Sci. 113:4121-4135(2000).
 CC -1- FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS AT ADHERENS
 CC JUNCTIONS (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 10 ARM REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AJ243418; -, NOT_ANNOTATED_CDS.

DR HSSP: P35222; IG3J.

DR MGD: MG1:109620; Arvcf.

DR InterPro: IPR000225; Armadillo.

DR Pfam: PF00514; Armadillo_seg; 6.

DR SMART: SM00185; ARM; 4.

DR PROSITE: PS50176; ARM_REPEAT; 3.

KW Cell adhesion; Cytoskeleton; Structural protein; Repeat; Coiled coil.

FT NON_TER

FT DOMAIN

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

FT REPEAT

SEQUENCE 969 AA; 105378 MW; 91130069E484C5B5 CRC64;

Query Match 49.4%; Score 38; DB 1; Length 969;

Best Local Similarity 61.5%; Pred. NO. 1.2e+02;

Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 NFAGIEAASAIQ 14

Db 664 NFNTLEAAGALQ 676

Search completed: July 3, 2003, 14:05:52
Job time: 7.5 secs

A;Residues: 1-714 <STO>

A:Cross-references: GB:AE005176; PID:g12724781; PIDN:AAK05857.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: glpP

Query Match 61.0%; Score 47; DB 2; Length 714;
Best Local Similarity 50.0%; Pred. No. 4.5;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 WNFAGIEAASAI0 14
Db 303 WNFIEGFAAVDSV0 316

RESULT 3

urease accessory protein D [imported] - Rhodobacter sphaeroides
C:Species: Rhodobacter sphaeroides
C>Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C:Accession: T50707
R:Choudhary, M.; Kaplan, S.
Nucleic Acids Res. 28, 862-867, 2000
A:Title: DNA sequence analysis of the photosynthesis region of Rhodobacter sphaeroides 2
A:Reference number: 225222; MUID:20115911; PMID:10648776
A:Accession: T50707
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-225 <CHO>
A:Cross-references: EMBL:AF195122; PIDN:AAF24251.1
A:Experimental source: strain 2.4.1
C:Genetics:
A:Gene: ured

Query Match 57.1%; Score 44; DB 2; Length 225;
Best Local Similarity 75.0%; Pred. No. 4.5;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 4 AGIEAASAI0G 15
Db 178 AGVEAASAFDG 189

RESULT 4

noea protein - Rhizobium meliloti
C:Species: Rhizobium meliloti
C>Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 04-Mar-2000
C:Accession: S71360
R:Ardourel, M.; Lortet, G.; Mallet, F.; Roche, P.; Truchet, G.; Prome, J.C.; Rosenberg,
submitted to the EMBL Data Library, May 1995
A:Reference number: S71360
A:Accession: S71360
A:Molecule type: DNA
A:Residues: 1-476 <ARD>
A:Cross-references: EMBL:U06430; NID:g1326068; PIDN:AAK44091.1; PID:g1326070
A:Experimental source: strain RCR2011
R:Ardourel, M.; Lortet, G.; Mallet, F.; Roche, P.; Truchet, G.; Prome, J.C.; Rosenberg,
Mol. Microbiol. 17, 687-699, 1995
A:Title: In Rhizobium meliloti, the operon associated with the nod box n5 comprises nodI
A:Reference number: S71357; MUID:96111489; PMID:8801423
A:Accession: S71358
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 304-339 <ARW>
A:Cross-references: EMBL:U06430
A:Experimental source: strain RCR2011 (-SU47)
C:Genetics:
A:Gene: noea
C:Superfamily: Rhizobium meliloti noea protein
C:Keywords: nodulation

Query Match 55.8%; Score 43; DB 2; Length 476;

Best Local Similarity 70.0%; Pred. No. 15;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 WNFAGIEAAM 10
Db 101 WSPFAGLKAAA 110

RESULT 5

Noea host specific nodulation protein [imported] - Sinorhizobium meliloti (strain 102
H95313
C:Species: Sinorhizobium meliloti
C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: H95313
R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; B
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meli
A:Reference number: A95262; MUID:21396509; PMID:11481432
A:Accession: H95313
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-476 <KUR>
A:Cross-references: GB:AE006469; PIDN:AAK65074.1; PID:g14523508; GSPDB:GN00165
A:Experimental source: strain 1021, megaplasmid pSymA
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubl
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau
hebaute, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yen,
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: noea
A:Genome: plasmid
C:Superfamily: Rhizobium meliloti noea protein

Query Match 55.8%; Score 43; DB 2; Length 476;
Best Local Similarity 70.0%; Pred. No. 15;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 WNFAGIEAAM 10
Db 101 WSPFAGLKAAA 110

RESULT 6

conserved hypothetical protein PA3332 [imported] - Pseudomonas aeruginosa (strain PAO
H83229
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83229
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pa
A:Reference number: A82950; MUID:20457337; PMID:10984043
A:Accession: H83229
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-141 <STO>
A:Cross-references: GB:AE004755; GB:AE004091; NID:g9949456; PIDN:AMG06720.1; GSPDB:GN
A:Experimental source: strain PAO1
C:Genetics:
A:Gene: PA3332
C:Superfamily: Bacillus subtilis hypothetical protein yese

Query Match 55.2%; Score 42.5; DB 2; Length 141;
Best Local Similarity 40.9%; Pred. No. 5.1;
Matches 9; Conservative 2; Mismatches 4; Indels 7; Gaps 1;

QY 1 WN-----FACIEAASAI0G 15
||| ||| ||| |||
Db 119 WNPJRHLEALGCGVEAAKIV0G 140

RESULT 7

Re-S oxidoreductase [imported] - Chlamydia pneumoniae (strain J138)
E86554
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C:Accession: E86554
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ise
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349; PMID:10871362
A:Accession: E86554
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-369 <STO>
A:Cross-references: GB:BA000008; NID:98978883; PIDN:BA98719.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: CPJ0513
C:Superfamily: hypothetical protein AF0390

Query Match 54.5%; Score 42; DB 2; Length 369;
Best Local Similarity 77.8%; Pred. No. 17;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WNFAGIEAA 9
||| ||| ||| |||
Db 305 WNYLGIEAA 313

RESULT 8

Conserved hypothetical protein CP0240 [imported] - Chlamydia pneumoniae (strains CWI
N:Alternate names: fe-s oxidoreductase
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C:Accession: G72069; F81599
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Ollinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606; PMID:10192388
A:Accession: G72069
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-369 <ARN>
A:Cross-references: GB:AE001636; GB:AE001363; NID:94376794; PIDN:AA018653.1; PID:9437680
A:Experimental source: strain CWI029
R:Read, T.D.; Brumham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255; PMID:10684935
A:Accession: F81599
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-369 <REA>
A:Cross-references: GB:AE002184; GB:AE002161; NID:97189156; PIDN:AAF8105.1; PID:9718916
C:Genetics:
A:Gene: CPN0513; CP0240
C:Superfamily: hypothetical protein AF0390

Query Match 54.5%; Score 42; DB 2; Length 369;
Best Local Similarity 77.8%; Pred. No. 17;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WNFAGIEAA 9
||| ||| ||| |||

Db 305 WNYLGIEAA 313

RESULT 9

ferrichrome ABC transporter (permease) fhvB - Bacillus subtilis
A69622
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C:Accession: A69622
R:Kunst, F.; Ogawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Acevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lechi, J.; Harwood, C.R.; Hentut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardio
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y. M.; Ogawa, K.; Ogilawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl
A:Authors: Schleib, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowaka, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchida
T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: A69622
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-384 <KUN>
A:Cross-references: GB:299121; GB:AL009126; NID:92635827; PIDN:CAB15337.1; PID:926358
A:Experimental source: strain 168
C:Genetics:
A:Gene: fhvB
C:Superfamily: vitamin B12 transport protein btuc

Query Match 53.2%; Score 41; DB 2; Length 384;
Best Local Similarity 46.7%; Pred. No. 26;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 WNFAGIEAASAI0G 15
||| ||| ||| |||
Db 174 WNFAGIEAASAI0G 188

RESULT 10

Probable cadaverine/lysine antiporter CadB VC0280 [imported] - Vibrio cholerae (strai
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: DB2342
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
Chaidson, D.; Ermolaeva, M.D.; Vamthanan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers
I. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: DB2342
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-445 <HEI>
A:Cross-references: GB:AE004116; GB:AE003852; NID:99654687; PIDN:AAF93455.1; GSPDB:GN
A:Experimental source: serogroup O1, strain N16961, biotype El Tor
C:Genetics:
A:Gene: VC0280
A:Map position: 1
C:Superfamily: L-lysine transport protein

Query Match 53.2%; Score 41; DB 2; Length 445;
Best Local Similarity 60.0%; Pred. No. 30;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 WNFAGIEAASAI 10
||| ||| ||| |||

Db 199 MSFVGVESA 208

RESULT 11

ABC transporter permease protein yj1D [Imported] - Lactococcus lactis subsp. lactis (str F86745)
 C:Species: Lactococcus lactis subsp. lactis
 C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
 C:Accession: F86745
 R:Holotin, A.; Winkler, P.; Mueger, S.; Jalllon, O.; Malarme, K.; Weissenbach, J.; Ehrlich Genome Res. 11, 731-753, 2001
 A>Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
 A:Reference number: AB6625; MUID:21235186; PMID:11337471
 A:Accession: F86745
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-540 <STO>
 A:Cross-references: GB:AE005176; PID:g12723906; PIDN:AAK05064.1; GSPDB:GN00146
 A:Experimental source: strain IL1403
 C:Genetics:
 A:Gene: yj1D

Query Match 53.2%; Score 41; DB 2; Length 540;
 Best Local Similarity 69.2%; Pred. No. 37;
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 3 FAGIEAASAI0G 15
 Db 339 FVGNETAASAI0G 351

RESULT 12

glyceraldehyde-3-phosphate dehydrogenase (phosphorylating) (EC 1.2.1.12) - Escherichia h
 C:Species: Escherichia blattae
 C:Date: 04-Sep-1997 #sequence_revision 07-Nov-1997 #text_change 03-Jun-2002
 C:Accession: I41220; I41223; I41226
 R:Lawrence, J.G.; Ochman, H.; Hartl, D.L.
 J. Gen. Microbiol. 137, 1911-1921, 1991
 A>Title: Molecular and evolutionary relationships among enteric bacteria.
 A:Reference number: I40701; MUID:92065252; PMID:1955870
 A:Accession: I41220
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-294 <RES>
 A:Cross-references: GB:M63358; NID:g146076; PIDN:AAA23837.1; PID:g146077
 A:Accession: I41223
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-294 <RES>
 A:Cross-references: GB:M63359; NID:g146108; PIDN:AAA23852.1; PID:g146109
 A:Experimental source: ATCC 33429
 A:Accession: I41226
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-294 <RES>
 A:Cross-references: GB:M63360; NID:g146114; PIDN:AAA23855.1; PID:g146115
 A:Experimental source: ATCC 33430
 C:Superfamily: glyceraldehyde-3-phosphate dehydrogenase
 C:Keywords: oxidoreductase

Query Match 51.9%; Score 40; DB 2; Length 294;
 Best Local Similarity 58.3%; Pred. No. 29;
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 WNFAGIEAASAI 12
 Db 70 WNEAGIEVVAEA 81

RESULT 13

B97457
 BH3305 conserved hypothetical [Imported] - Agrobacterium tumefaciens (strain C58, Cereon

C:Species: Agrobacterium tumefaciens
 C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
 C:Accession: B97457

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirillo, B.; Goldm
 A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz,
 Science 294, 2323-2328, 2001
 A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
 A:Reference number: A97359; PMID:11743194
 A:Accession: B97457
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-323 <KUR>
 A:Cross-references: GB:AE007869; PIDN:AAK86611.1; PID:g15155783; GSPDB:GN00169
 C:Genetics:
 A:Gene: AGR_C1468
 A:Map position: circular chromosome

Query Match 51.9%; Score 40; DB 2; Length 323;
 Best Local Similarity 61.5%; Pred. No. 32;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 3 FAGIEAASAI0G 15
 Db 11 FAGISCAATPALAG 23

RESULT 14

conserved hypothetical protein Atu0802 [Imported] - Agrobacterium tumefaciens (strain
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
 C:Accession: AD2675
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
 erage, G.; Giallet, W.; Grant, C.; Guenther, D.; Kuyavin, T.; Levy, R.; Li, M.; McCl
 science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
 ster, E.W.
 A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; PMID:11743193
 A:Accession: AD2675
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-323 <KUR>
 A:Cross-references: GB:AE008688; PIDN:AAI41818.1; PID:g17739175; GSPDB:GN00186
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: Atu0802
 A:Map position: circular chromosome

Query Match 51.9%; Score 40; DB 2; Length 323;
 Best Local Similarity 61.5%; Pred. No. 32;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 3 FAGIEAASAI0G 15
 Db 11 FAGISCAATPALAG 23

RESULT 15

probable Na⁺/Ca²⁺-K⁺-exchanging protein - Methanobacterium thermoautotrophicum (strat
 C:Species: Methanobacterium thermoautotrophicum
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
 C:Accession: H69020
 R:Smith, D.R.; Doucette-Stamm, L.A.; Delonghery, C.; Lee, H.; Dubois, J.; Aldredge, T
 ; Qiu, D.; Spadofora, R.; Vitale, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani,
 K.I.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
 J. Bacteriol. 179, 7135-7155, 1997
 A>Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
 A:Reference number: A69000; MUID:98037514; PMID:9371463
 A:Accession: H69020
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
A:Residues: 1-330 <MTH>
A:Cross-references: GB:AE000885; GB:AE000666; NID:g2622256; PIDN:AA885644.1; PID:g262226
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH155
A:Start codon: TTG

Query Match 51.9%; Score 40; DB 2; Length 330;
Best Local Similarity 46.7%; Pred. No. 33;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 WNFAGIEPAASAIQG 15
||| | : | : |
DB 95 WNIAGILGISSVAG 109

Search completed: July 3, 2003, 14:07:19
Job time : 13.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2003, 14:00:24 ; Search time 30 Seconds

(without alignments)
66.625 Million cell updates/sec

Title: US-09-830-839-6

Perfect score: 77

Sequence: 1 MNFAGIEAASAIQ 15

Scoring table: BLOSUM62

Gapped 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A.GeneSeq_101002:*
1: /SIDSz/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT:*
2: /SIDSz/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT:*
3: /SIDSz/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT:*
4: /SIDSz/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT:*
5: /SIDSz/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT:*
6: /SIDSz/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT:*
7: /SIDSz/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT:*
8: /SIDSz/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT:*
9: /SIDSz/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT:*
10: /SIDSz/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT:*
11: /SIDSz/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT:*
12: /SIDSz/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT:*
13: /SIDSz/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT:*
14: /SIDSz/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT:*
15: /SIDSz/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT:*
16: /SIDSz/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT:*
17: /SIDSz/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT:*
18: /SIDSz/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT:*
19: /SIDSz/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT:*
20: /SIDSz/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT:*
21: /SIDSz/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT:*
22: /SIDSz/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT:*
23: /SIDSz/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 77 | 100.0 | 15 | AA194595 | Mycobacterium tube |
| 2 | 77 | 100.0 | 15 | AAO17433 | Early secreted ant |
| 3 | 77 | 100.0 | 20 | AAE12275 | Mycobacterium tube |
| 4 | 77 | 100.0 | 51 | AAW32466 | Mycobacterium tube |
| 5 | 77 | 100.0 | 51 | AAW32339 | Mycobacterium tube |
| 6 | 77 | 100.0 | 51 | AAW81701 | M. tuberculosis im |
| 7 | 77 | 100.0 | 51 | AAW64334 | Mycobacterium tube |
| 8 | 77 | 100.0 | 51 | AAV39131 | M. tuberculosis ES |
| 9 | 77 | 100.0 | 51 | AAV38988 | M. tuberculosis re |
| 10 | 77 | 100.0 | 22 | AAU01897 | Mycobacterium tube |

| | | | | | |
|----|----|-------|------|----------|---------------------|
| 11 | 77 | 100.0 | 51 | AAE17580 | Mycobacterium spec |
| 12 | 77 | 100.0 | 95 | AAW11494 | ESAT-6, Mycobacter |
| 13 | 77 | 100.0 | 95 | AAV29890 | Mycobacterium tube |
| 14 | 77 | 100.0 | 95 | AAV29788 | Mycobacterium tube |
| 15 | 77 | 100.0 | 95 | AAV35219 | M. tuberculosis RV3 |
| 16 | 77 | 100.0 | 95 | AAW50740 | Mycobacterium tube |
| 17 | 77 | 100.0 | 196 | AAV29888 | Mycobacterium tube |
| 18 | 77 | 100.0 | 403 | AAV29843 | Mycobacterium tube |
| 19 | 77 | 100.0 | 403 | AAV21963 | Amino acid sequenc |
| 20 | 77 | 100.0 | 404 | AAW72842 | Mycobacterium tube |
| 21 | 77 | 100.0 | 404 | AAV21962 | Amino acid sequenc |
| 22 | 77 | 92.2 | 19 | AAW35548 | ESAT-6 antigen SEQ |
| 23 | 66 | 85.7 | 20 | AAE12276 | Mycobacterium tube |
| 24 | 62 | 80.5 | 17 | AAV88581 | Antigenic N-termi |
| 25 | 60 | 77.9 | 13 | AAE12274 | Mycobacterium tube |
| 26 | 54 | 70.1 | 15 | AAV94590 | Mycobacterium tube |
| 27 | 54 | 70.1 | 15 | AAO17432 | Early secreted ant |
| 28 | 47 | 61.0 | 714 | ABW55108 | Mycobacterium tube |
| 29 | 44 | 57.1 | 15 | AAV94599 | Mycobacterium tube |
| 30 | 42 | 54.5 | 409 | AAV35136 | Chlamydia pneumoni |
| 31 | 41 | 53.2 | 540 | ABW54282 | Lactococcus lactis |
| 32 | 40 | 51.9 | 81 | AAU42889 | Propionilbacterium |
| 33 | 40 | 51.9 | 403 | AAW40540 | Human polyprotein |
| 34 | 40 | 51.9 | 2293 | AAV69197 | Amino acid sequenc |
| 35 | 39 | 50.6 | 9 | AAW61077 | Amino acid sequenc |
| 36 | 39 | 50.6 | 140 | AAW41828 | Human OREX ORF1592 |
| 37 | 39 | 50.6 | 162 | AAW40271 | Human polyprotein |
| 38 | 39 | 50.6 | 162 | AAW93538 | Human protein sequ |
| 39 | 39 | 50.6 | 162 | AAV72606 | Human Electron tra |
| 40 | 39 | 50.6 | 209 | AAW36643 | Arabidopsis thalia |
| 41 | 39 | 50.6 | 221 | AAV91434 | Human secreted pro |
| 42 | 39 | 50.6 | 228 | AAW42057 | Human polyprotein |
| 43 | 39 | 50.6 | 232 | AAW36642 | Arabidopsis thalia |
| 44 | 39 | 50.6 | 232 | AAU21824 | Novel human neopla |
| 45 | 39 | 50.6 | 253 | AAW23550 | Human EST encoded |

ALIGNMENTS

RESULT 1
AA194595
ID AAY94595 standard; peptide; 15 AA.
XX
AC AAY94595;
DT 28-NOV-2000 (first entry)
XX
DE Mycobacterium tuberculosis ESAT-6 peptide ES6.
XX
KW Tuberculosis; infection diagnosis; ESAT-6.
XX
OS Mycobacterium tuberculosis.
XX
PN WO200026248-A2.
PD 11-MAY-2000.
XX
PF 03-NOV-1999; 99MO-GB03635.
XX
PR 04-NOV-1998; 98GB-0024213.
PR 04-NOV-1998; 98US-0107004.
XX
PA (ISIS-) ISIS INNOVATION LTD.
XX
PI Lalvani A, Pathan AA;
XX
DR WPI; 2000-365579/31.
XX
PT Novel method of diagnosing infection, or exposure of a host, to a
PT Mycobacterium comprising contacting T cells from the host with ESAT-6
PT derived peptides -

PS Claim 1; Page 3; 33pp; English.
XX
CC The present sequence is the peptide ES6 derived from the Mycobacterium
CC tuberculosis ESAT-6 gene. This sequence is one of eleven peptides
CC derived from the ESAT-6 gene (see AAY94590 to AAY94600). The peptides
CC are recognized, to varying degrees, by the T cells of patients with
CC tuberculosis. When the peptides are combined together as a panel they
CC provide a high specificity and sensitivity diagnostic test for
CC M. tuberculosis infection. This test has the advantage that it does not
CC give false positives for patients who have been vaccinated with BCG.
XX
SQ Sequence 15 AA:
Query Match 100.0%; Score 77; DB 21; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.4e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 WNFAGIEAASAIQG 15
DB 1 WNFAGIEAASAIQG 15
RESULT 2
AAO17433
ID AAO17433 standard; peptide; 15 AA.
XX
AC AAO17433;
XX
DT 11-JUL-2002 (first entry)
XX
DE Early secreted antigenic target 6 T cell epitope #2.
XX
XX Tuberculosis; TB; vaccination; vaccine; CD4+ T cell immune response;
KM poxvirus vector; HIV; malaria; Helicobacter pylori; influenza;
KM hepatitis; viral infection; leprosy; protozoan parasite; cancer;
KM tuberculosis; anti-HIV; protozoicide; antibacterial; vituicide;
KM hepatocytic; anti-inflammatory; antileptotic; cytoskeletal; epitope.
XX
OS Mycobacterium tuberculosis.
XX
PN WO200224224-A2.
XX
PD 28-MAR-2002.
XX
PF 13-SEP-2001; 2001WO-GB04116.
XX
PR 21-SEP-2000; 2000GB-0023203.
XX
PA (OXO-) OXON PHARMACEUTICALS LTD.
XX
PI Hill AVS, McShane H, Gilbert S, Reece W, Schneider J;
XX
DR WPI; 2002-394098/42.
XX
PT Inducing CD4+ T-cell response against target antigen by administering a
PT composition comprising a source of CD4+ epitopes which is a
PT non-replicating or replication impaired recombinant poxvirus vector
XX
PS Example 1; Page 21; 50pp; English.
XX
CC The present invention relates to a method of inducing a CD4+ T-cell
CC response against a target antigen, by administering two different
CC compositions comprising a source of CD4+ T-cell epitope(s) of the target
CC antigen, where the second composition further includes an epitope which
CC is same as the epitope of the first composition, where the source of the
CC epitopes for the compositions is a non-replicating or replication
CC impaired recombinant poxvirus vector. The methods are useful for inducing
CC CD4+ T-cell immune responses against diseases such as tuberculosis, human
CC immunodeficiency virus (HIV) (persistent viral infection), malaria,
CC Helicobacter pylori, influenza, hepatitis (chronic hepatitis B and C),
CC cytomegalovirus (CMV), viral infection, herpes virus-induced disease,
CC leprosy and diseases caused by non-malarial protozoan parasite such as
CC toxoplasma and cancer. The present sequence is an epitope from

CC Mycobacterium tuberculosis useful in the method of the invention.
XX
SQ Sequence 15 AA;
Query Match 100.0%; Score 77; DB 23; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.4e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 WNFAGIEAASAIQG 15
DB 1 WNFAGIEAASAIQG 15
RESULT 3
AAE12275
ID AAE12275 standard; peptide; 20 AA.
XX
AC AAE12275;
XX
DT 18-DEC-2001 (first entry)
XX
DE Mycobacterium tuberculosis (Mtb) peptide #60.
XX
XX Mycobacterium tuberculosis; Mtb peptide; antibacterial; vaccine;
KM infection; anti-Mtb immune response.
XX
OS Mycobacterium tuberculosis.
XX
PN WO200170774-A2.
XX
PD 27-SEP-2001.
XX
PF 20-MAR-2001; 2001WO-US08906.
XX
PR 20-MAR-2000; 2000US-190834P.
XX
PA (UYBR-) UNIV BROWN RES. FOUND.
XX
PI Degroot AS;
XX
DR WPI; 2001-616401/71.
XX
PT New vaccine for immunizing a mammalian subject, preferably humans,
PT against infection caused by Mycobacterium tuberculosis
XX
PS Disclosure; Fig 4; 42pp; English.
XX
CC The present invention relates to Mycobacterium tuberculosis (Mtb)
CC vaccine candidate peptides. The invention also relates to a method
CC for identifying Mtb vaccine candidate peptides as well as vaccines
CC comprising these candidate peptides. Vaccines of the invention and
CC Mtb vaccine candidate peptides are useful for inducing an anti-
CC Mycobacterium tuberculosis (anti-Mtb) immune response by raising
CC anti-Mtb antibody in a mammalian subject preferably human. They
CC are used for immunizing a mammalian subject, preferably humans,
CC against infection caused by Mycobacterium tuberculosis. The present
CC sequence is a Mtb vaccine candidate peptide.
XX
SQ Sequence 20 AA;
Query Match 100.0%; Score 77; DB 22; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.9e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 WNFAGIEAASAIQG 15
DB 2 WNFAGIEAASAIQG 16
RESULT 4
AAW32466
ID AAW32466 standard; Protein; 51 AA.
XX

```

AC AAW32466;
XX
DT 09-JAN-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen ESAT-6.
XX
KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
XX skin testing; M.tuberculosis.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9709428-A2.
XX
PD 13-MAR-1997.
XX
PF 30-AUG-1996; 96WO-US14674.
XX
PR 12-JUL-1996; 96US-0680574.
XX 01-SEP-1995; 95US-0523436.
XX 22-SEP-1995; 95US-0533634.
XX 22-MAR-1996; 96US-0620874.
XX 05-JUN-1996; 96US-0659683.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAM;
PI Twardzik DR, Vedyick TH;
XX
DR WPI; 1997-192903/17.
XX N-PSDB; AAT91529.
XX
PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
PT useful in vaccines for prevention or treatment of tuberculosis, also
PT for diagnosis
XX
PS Disclosure; Page 135; 168pp; English.
XX
CC A new immunogenic polypeptide has been developed comprising an
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
CC its variant differing only in conservative substitutions and/or
CC modifications). The present sequence represents a M.tuberculosis
CC antigen, ESAT-6. The immunogenic protein, and fusion proteins
CC containing one or more of the proteins or one of the proteins plus
CC ESAT-6, are useful in vaccines, preferably when formulated with a
CC non-specific adjuvant, to induce an immune response against
CC M.tuberculosis (for treatment or prevention).
XX
SQ Sequence 51 AA;
Query Match 100.0%; Score 77; DB 18; Length 51;
Best Local Similarity 100.0%; Pred. No. 5.2e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 WNFAGIEAASAIOG 15
DB 6 WNFAGIEAASAIOG 20
RESULT 5
AAW32339
ID AAW32339 standard; Protein; 51 AA.
XX
AC AAW32339;
XX
DT 13-JAN-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen ESAT-6.
XX
KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
XX skin testing; M.tuberculosis.
XX
OS Mycobacterium tuberculosis.
XX

```

```

PN WO9709429-A2.
XX
PD 13-MAR-1997.
XX
PF 30-AUG-1996; 96WO-US14675.
XX
PR 12-JUL-1996; 96US-0680573.
XX 01-SEP-1995; 95US-0523435.
XX 22-SEP-1995; 95US-0532136.
XX 22-MAR-1996; 96US-0620280.
XX 05-JUN-1996; 96US-0658800.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAM;
PI Twardzik DR, Vedyick TH;
XX
DR WPI; 1997-192904/17.
XX N-PSDB; AAT91463.
XX
PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens
PT - useful for diagnosis of M. tuberculosis infection
XX
PS Claim 43; Page 147; 190pp; English.
XX
CC A new immunogenic polypeptide has been developed comprising an
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
CC its variant differing only in conservative substitutions and/or
CC modifications). The present sequence represents a specifically claimed
CC sequence of the ESAT-6 M.tuberculosis antigen. The immunogenic
CC polypeptide can be used to diagnose M.tuberculosis infection by forming
CC complexes with specific antibodies in the sample. Fragments of DNA
CC encoding the immunogenic polypeptide can be used as diagnostic primers
CC or probes and agents that bind to the antigen, especially monoclonal
CC antibodies or equivalent polyclonal antibodies, are also used for
CC diagnosis.
XX
SQ Sequence 51 AA;
Query Match 100.0%; Score 77; DB 18; Length 51;
Best Local Similarity 100.0%; Pred. No. 5.2e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 WNFAGIEAASAIOG 15
DB 6 WNFAGIEAASAIOG 20
RESULT 6
AAW81701
ID AAW81701 standard; Protein; 51 AA.
XX
AC AAW81701;
XX
DT 27-JAN-1999 (first entry)
XX
DE M. tuberculosis immunogenic polypeptide ESAT-6.
XX
KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
KW vaccine; pharmaceutical; infection; diagnosis.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9816646-A2.
XX
PD 23-APR-1998.
XX
PF 07-OCT-1997; 97WO-US18293.
XX 13-MAR-1997; 97US-0818112.
XX 11-OCT-1996; 96US-0730510.
XX
PA (CORI-) CORIXA CORP.

```

XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
 PI Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;
 XX WPI: 1998-261042/23.
 DR N-PSDB: AAV64501.
 XX
 PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
 PT to develop products for the detection of M. tuberculosis infection
 PT and for diagnosis, treatment and prevention of tuberculosis
 XX
 PS Disclosure; Page 126; 230pp; English.
 XX
 CC This sequence represents an immunogenic portion of a soluble
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method
 CC for inducing protective immunity against tuberculosis (TB). This
 CC sequence can be formulated into vaccines and/or pharmaceutical
 CC compositions for immunising against M. tuberculosis infection or may
 CC be used for the diagnosis of tuberculosis.
 CC
 SQ Sequence 51 AA;

Query Match 100.0%; Score 77; DB 19; Length 51;
 Best Local Similarity 100.0%; Pred. No. 5.2e-06;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WNFAGIEAASAI0G 15
 |||||
 Db 6 WNFAGIEAASAI0G 20

RESULT 7
 AAW64334
 ID AAW64334 standard; Protein; 51 AA.
 XX
 AC AAW64334;
 XX
 DT 09-NOV-1998 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen ESAT-6.
 XX
 KM Tuberculosis; infection; diagnosis; antigen; ESAT-6.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9816645-A2.
 XX
 PD 23-APR-1998.
 XX
 PF 07-OCT-1997; 97WO-US18214.
 XX
 PR 13-MAR-1997; 97US-0818111.
 PR 11-OCT-1996; 96US-0729622.
 XX
 PA (CORT-) CORIXA CORP.
 XX
 PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
 PI Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;
 XX
 DR WPI: 1998-251292/22.
 DR N-PSDB: AAV44393.
 XX
 PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used
 PT to develop products for the detection of M. tuberculosis infection
 PT and diagnosis of tuberculosis
 XX
 PS Disclosure; Page 130; 250pp; English.
 XX
 CC This polypeptide comprises an antigenic portion of Mycobacterium
 CC tuberculosis antigen ESAT-16. The invention relates to
 CC compositions and methods for diagnosing tuberculosis. It provides
 CC polypeptides (see AAW64291-W64379) comprising an antigenic portion
 CC of a soluble M. tuberculosis antigen, or an immunogenic portion of

CC an M. tuberculosis antigen, as well as fusion proteins between
 CC these polypeptides and known antigens such as ESAT-6. Also claimed
 CC are methods and diagnostic kits for detecting M. tuberculosis
 CC infection in a patient using these polypeptides, antibodies, or
 CC oligonucleotide probes and primers.
 CC
 SQ Sequence 51 AA;

Query Match 100.0%; Score 77; DB 19; Length 51;
 Best Local Similarity 100.0%; Pred. No. 5.2e-06;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WNFAGIEAASAI0G 15
 |||||
 Db 6 WNFAGIEAASAI0G 20

RESULT 8
 AAY39131
 ID AAY39131 standard; Protein; 51 AA.
 XX
 AC AAY39131;
 XX
 DT 05-NOV-1999 (first entry)
 XX
 DE M. tuberculosis ESAT-6 amino acid sequence.
 XX
 KM Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
 KM immunotherapy; diagnosis; immunisation; vaccine; infection;
 KM immune response; skin test.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9942076-A2.
 XX
 PD 26-AUG-1999.
 XX
 PF 17-FEB-1999; 99WO-US03268.
 XX
 PR 05-MAY-1998; 98US-0072967.
 PR 18-FEB-1998; 98US-0025197.
 XX
 PA (CORT-) CORIXA CORP.
 XX
 PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;
 XX
 DR WPI: 1999-527409/44.
 DR N-PSDB: AA219303.
 XX
 PT New antigens from Mycobacterium tuberculosis useful in diagnostic
 PT skin tests and protective or therapeutic vaccines or compositions
 XX
 PS Disclosure; Page 120; 299pp; English.
 XX
 CC The present invention describes polypeptides comprising an immunogenic
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
 CC are vaccines and fusion protein containing M. tuberculosis Ag's.
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
 CC other polypeptides fragments, can be used in pharmaceutical compositions
 CC or vaccines to generate a protective or therapeutic immune response to
 CC M. tuberculosis and as reagents in skin tests for diagnosis of
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
 CC by, T, B or natural killer cells and/or macrophages in
 CC tuberculosis-immune subjects. AA219249 to AA219460 and AAY39083 to
 CC AAY39225 are used in the exemplification of the present invention.
 CC
 SQ Sequence 51 AA;

Query Match 100.0%; Score 77; DB 20; Length 51;
 Best Local Similarity 100.0%; Pred. No. 5.2e-06;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNFAGIEAASAI0G 15
 |||||
 DB 6 MNFAGIEAASAI0G 20

RESULT 9

AA38988
 ID AAY38988 standard; Protein: 51 AA.

AC AAY38988;

DT 05-NOV-1999 (first entry)

DE M. tuberculosis recombinant antigen protein ESAT-6.

KW Antigen; diagnosis; detection; infection; antibody; immunisation;
 vaccine; immunity.

OS Mycobacterium tuberculosis.

PN W09942118-A2.

PD 26-AUG-1999.

PF 17-FEB-1999; 99WO-US03265.

PR 05-MAY-1998; 98US-0072596.

PR 18-FEB-1998; 98US-0024753.

PA (CORI-) CORIXA CORP.

PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 Lodes MJ, Reed SG, Skeiky YAM, Twardzik DR, Vedyck TS;

DR WPI: 1999-527416/44.

DR N-PSDB: AA219091.

PT New polypeptide comprising antigenic portions of M. tuberculosis

PS Claim 51; Page 165; 323pp; English.

CC This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against
 CC M. tuberculosis infection. The new detection methods are needed as
 CC current vaccination strategies do not provide 100% immunity.

SO Sequence 51 AA;

Query Match 100.0%; Score 77; DB 20; Length 51;

Best Local Similarity 100.0%; Pred. No. 5.2e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNFAGIEAASAI0G 15
 |||||
 DB 6 MNFAGIEAASAI0G 20

RESULT 10

AAU01897
 ID AAU01897 standard; Protein: 51 AA.

AC AAU01897;

DT 29-AUG-2001 (first entry)

DE Mycobacterium tuberculosis partial antigen ESAT-6.

KW ESAT-6; antigen; vaccine; tuberculosis; AIDS;
 acquired immunodeficiency disease.

OS Mycobacterium tuberculosis.

PN W0200124820-A1.

PD 12-APR-2001.

PF 10-OCT-2000; 2000WO-US28095.

PR 07-OCT-1999; 99US-0158338.

PR 07-OCT-1999; 99US-0158425.

PA (CORI-) CORIXA CORP.

PI Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes MJ;

DR WPI: 2001-290576/30.

DR N-PSDB: AA503787.

PT Vaccinating against Mycobacteria infections in mammals using fusion
 proteins comprising combinations of heterologous antigens

PS Disclosure: Page 164; 168pp; English.

CC The sequence represents Mycobacterium tuberculosis ESAT-6 antigen.
 CC Compositions comprising at least 2 heterologous antigens, as a fusion
 CC protein, and vectors expressing the fusion proteins are used as vaccines
 CC to prophylactically immunise mammals (especially humans) against
 CC infection by Mycobacteria. The compositions contain at least 2
 CC heterologous antigens that increase the serological sensitivity of
 CC individuals infected with tuberculosis, a disease frequently affecting
 CC patients with acquired immunodeficiency disease, AIDS.

SO Sequence 51 AA;

Query Match 100.0%; Score 77; DB 22; Length 51;

Best Local Similarity 100.0%; Pred. No. 5.2e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNFAGIEAASAI0G 15
 |||||
 DB 6 MNFAGIEAASAI0G 20

RESULT 11

AAE17580
 ID AAE17580 standard; Protein: 51 AA.

AC AAE17580;

DT 22-APR-2002 (first entry)

DE Mycobacterium species ESAT-6 protein.

KW Fusion protein; antigen; serological sensitivity; Immune response;
 tuberculosis; infection; vaccine; ESAT-6 protein.

OS Mycobacterium sp.

PN W0200198460-A2.

PD 27-DEC-2001.

PF 20-JUN-2001; 2001WO-US19959.

PR 20-JUN-2000; 2000US-0597796.

PR 01-FEB-2001; 2001US-265737P.

PA (CORI-) CORIXA CORP.

PI Skeiky Y, Reed S, Alderson M;

DR WPI: 2002-147798/19.
 DR N-PSDB: AAD28351.

XX Composition comprising MTB39 antigen and MTB32A antigen from
PT Mycobacterium species, useful for eliciting immune response in a
PT subject
XX
XX Claim 9: Page 122; 136pp; English.
XX
CC The present invention relates to fusion proteins containing at least
CC two Mycobacterium species antigens, nucleotides encoding them and
CC compositions comprising such fusion proteins. The present invention
CC particularly relates to nucleic acids encoding fusion proteins that
CC include two or more individual M. tuberculosis antigens which increase
CC the serological sensitivity of sera from individuals infected with
CC tuberculosis and methods for their use in diagnosis, prevention and
CC treatment of tuberculosis infection. Sequences of the invention are
CC useful for eliciting an immune response in a mammal, e.g., human,
CC immunised with BCG. They are useful in the diagnosis, treatment and
CC prevention of Mycobacterium infection. The fusion proteins and the
CC polynucleotides are useful as diagnostic tools in patients infected
CC with Mycobacterium, in vitro and in vivo assays for detecting humoral
CC antibodies or cell-mediated immunity against M. tuberculosis, for the
CC diagnosis of an infection or monitoring of disease progression, as
CC immunogens to generate or elicit a protective immune response in a
CC patient and for raising anti-M. tuberculosis antibodies in a non-human
CC animal. Sequences of the invention are also used as vaccines. MTB32A
CC fusion proteins of the invention are useful as in vivo diagnostic agents
CC for intradermal skin test. The present sequence is Mycobacterium species
CC ESAT-6 protein.
CC
XX
SQ Sequence 51 AA:

Query Match 100.0%; Score 77; DB 23; Length 51;
Best Local Similarity 100.0%; Pred. No. 5.2e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNFAGIEAASAIQG 15
| | | | | | | | | | | | | | | | |
DB 6 MNFAGIEAASAIQG 20

RESULT 12
AAW11494
ID AAW11494 standard; Protein; 95 AA.
XX
AC AAW11494;
XX
DT 27-MAR-1997 (first entry)
XX
DE ESAT6.
XX
KM ESAT6; Mycobacterium tuberculosis; 6kDa antigen; tuberculosis complex;
KM HXB76-8 reactive antigen; interferon-gamma; memory T-lymphocyte; vaccine;
XX M. africanum; M. bovis; delayed-type hypersensitivity reaction; therapy.
XX
OS Mycobacterium tuberculosis.
XX
PN W09501441-A1.
XX
PD 12-JAN-1995.
XX
PF 01-JUL-1994; 94MO-DK00273.
XX
PR 02-JUL-1993; 93DK-0000798.
XX
PA (STAT-) STATENS SERUMSINSTITUT.
XX
PI Andersen AB, Andersen P, Haslov K, Sorensen A;
XX
DR WPI; 1995-061005/08.
DR N-PSDB; AAT51422.
XX
PT Vaccine for tuberculosis induces interferon-gamma release from
T-lymphocytes - comprises an antigen released from mycobacteria,

PT for immunisation of humans
XX
XX Claim 5; Page 61-63; 101pp; English.
XX
CC This sequence represents the Mycobacterium tuberculosis ESAT6. ESAT6 is
CC also known as the 6kDa antigen, or the HXB76-8 reactive antigen. ESAT6
CC is released from metabolising bacteria, and can be isolated from short
CC term filtrates grown as shaken cultures for 7 days. ESAT6 also induces a
CC release of interferon-gamma from reactivated memory T-lymphocytes. This
CC protein sequence is included in the vaccine of the invention. The
CC vaccine is for immunising an animal (including humans) against
CC tuberculosis caused by a Mycobacteria belonging to the tuberculosis
CC complex. The Mycobacteria of the tuberculosis complex are, M.
CC tuberculosis, M. africanum, and M. bovis. The vaccine can evoke a
CC protective immune response against tuberculosis or a delayed-type
CC hypersensitivity reaction. The protein can also be included in a
CC composition for diagnosing tuberculosis. The composition is injected
CC intradermally, and a skin reaction is an indicator of tuberculosis.
CC
XX
SQ Sequence 95 AA:

Query Match 100.0%; Score 77; DB 16; Length 95;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNFAGIEAASAIQG 15
| | | | | | | | | | | | | | | | |
DB 6 MNFAGIEAASAIQG 20

RESULT 13
AAV29890
ID AAV29890 standard; Protein; 95 AA.
XX
AC AAV29890;
XX
DT 18-NOV-1999 (first entry)
XX
DE Mycobacterium tuberculosis ESAT-6 protein sequence.
XX
KM Mycobacterium tuberculosis; diagnosis; skin test; vaccine;
KM delayed type hypersensitivity; DTH; ESAT-6 homodimer; tuberculosis;
XX interferon-gamma release.
XX
OS Mycobacterium tuberculosis.
XX
PN W09945119-A2.
XX
PD 10-SEP-1999.
XX
PF 05-MAR-1999; 99MO-DK00109.
XX
PR 06-MAR-1998; 98DK-0000306.
XX
PR 06-MAR-1998; 98US-0077105.
XX
PA (STAT-) STATENS SERUM INST.
XX
PI Jensen CL, Folkersen J;
XX
DR WPI; 1999-551043/46.
DR N-PSDB; AA221132.
XX
PT New mycobacterial polypeptide produced in lactic acid bacteria, useful
in tuberculosis diagnosis and vaccines
XX
XX Disclosure; Page 75-76; 76pp; English.
XX
CC The present invention describes a bioactive polypeptide (or
CC immunologically equivalent analogue) produced in lactic acid bacteria
CC which reacts with lymphoid cells primed with Mycobacterium tuberculosis
CC complex mycobacteria (M. tuberculosis, M. africanum or M. bovis). The
CC polypeptide and ESAT-6 polypeptides are useful in compositions for
CC diagnosis of and vaccination against tuberculosis caused by

CC M. tuberculosis complex mycobacteria. The ESAT-6 polypeptide can be used
CC to diagnose ongoing/previous sensitization with these bacteria by
CC detecting cytokine release when contacting blood samples with the
CC polypeptide. The bioactive polypeptide may be used in diagnostic
CC compositions and vaccines for mycobacteria other than of the
CC M. tuberculosis complex, e.g. M. avium which infects poultry and
CC occasionally humans, M. leprae; they are especially useful when they do
CC not react with lymphoid cells previously primed with M. tuberculosis
CC complex mycobacteria, and so do not give rise to a diagnostic reaction
CC in individuals infected with these bacteria. The polypeptides may also
CC be used in in vitro diagnostic tests e.g. stimulation of interferon-gamma
CC release from lymphocytes. The polypeptide has similar or higher
CC bioactivity as currently used tuberculin reagent in the standard
CC delayed type hypersensitivity (DTH) skin test for tuberculosis, but may
CC have greater specificity, being better able to discriminate between
CC lymphoid cells primed from tuberculosis and from previous vaccination.
CC The present sequence represents M. tuberculosis ESAT-6 used in the
CC exemplification of the present invention.

CC Sequence 95 AA:

Query Match 100.0%; Score 77; DB 20; Length 95;

Best Local Similarity 100.0%; Pred. No. 1e-05; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WNFAGIEAASAIQG 15
|||||
DB 6 WNFAGIEAASAIQG 20

RESULT 14

AAY29788 standard; Protein; 95 AA.

AC AAY29788;

DT 08-NOV-1999 (first entry)

DE Mycobacterial tuberculosis ESAT-6 protein.

KW Mycobacterial; ESAT-6; 6 kDa antigen; tuberculosis; infection;

KW Immunological response; diagnosis; vaccine.

OS Mycobacterium tuberculosis.

PN US595077-A.

PD 21-SEP-1999.

PF 05-JUN-1995; 95US-0465640.

PR 05-JUN-1995; 95US-0465640.

PR 20-SEP-1993; 93US-0123182.

PR 01-JUL-1994; 94WO-DK00273.

PA (STAT-) STATENS SERUMINSTITUT.

PI Andersen AB, Andersen P, Haslov K, Sorensen AL;

DR WPI; 1999-539545/45.

DR N-PSDB; AAZ08877.

PT Polypeptide secreted from Mycobacterium is useful as a vaccine

PT against tuberculosis

PS Claim 24; Fig 10C; 39pp; English.

CC The present invention describes a purified or non-naturally occurring
CC polypeptide (I) released from a metabolising mycobacteria comprising an
CC ESAT-6, also called the 6 kDa antigen. The present sequence represents a
CC Mycobacterium tuberculosis ESAT-6 protein. Also described is a purified
CC or non-naturally occurring polypeptide (II) with the ability to elicit
CC a delayed type hypersensitivity reaction which comprises a T cell

CC epitope of (I). (II) can be used with a carrier or vehicle in a
CC composition for diagnosing tuberculosis caused by mycobacteria belonging
CC to the tuberculosis complex, i.e. Mycobacterium tuberculosis, M. bovis
CC and M. africanum. The composition can be used to detect microbial
CC antibodies or components of mycobacteria in samples or in animals
CC through the use of immunoassays. (II) can be used as a vaccine for
CC immunizing an animal, including humans against tuberculosis caused by
CC mycobacteria of the tuberculosis-complex. (II) induce a release of
CC IFN-gamma from reactivated T-lymphocytes evoking a protective immune
CC response. Vaccine containing (II) has the same protective potency as
CC the live BCG vaccine against tuberculosis.

CC Sequence 95 AA:

Query Match 100.0%; Score 77; DB 20; Length 95;

Best Local Similarity 100.0%; Pred. No. 1e-05; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WNFAGIEAASAIQG 15
|||||
DB 6 WNFAGIEAASAIQG 20

RESULT 15

AAB35219 standard; Protein; 95 AA.

AC AAB35219;

DT 24-APR-2001 (first entry)

DE M tuberculosis RV3875 protein.

XX

KW Tuberculosis; TB; vaccine; esat-6 gene family; RV0287; RV1036C;

KW RV1037C; RV2346C; RV2348C; RV2653C; RV2653C; RV3020C; RV3444C;

KW RV3445C; RV3890C; RV3891C; RV3904C; RV3905C.

XX

OS Mycobacterium tuberculosis.

PN WO200104151-A2.

PD 18-JAN-2001.

PF 13-JUL-2000; 2000WO-DK00398.

PR 13-JUL-1999; 99DK-0001020.

PR 15-JUL-1999; 99US-0144011.

PA (STAT-) STATENS SERUM INST.

PI Andersen P, Skjot R;

DR WPI; 2001-091923/10.

DR

PT New polypeptide encoded by a member of the esat-6-gene family for

PT immunizing against and diagnosis of tuberculosis -

PS Example 2; Page 65; 80pp; English.

CC The present invention provides the protein and coding sequences for
CC members of the esat-6 gene family from Mycobacterium tuberculosis. These
CC proteins include RV0287, RV1037C, RV2346C, RV2348C, RV2653C,
CC RV2654C, RV3020C, RV3444C, RV3445C, RV3890C, RV3891C, RV3904C and
CC RV3905C. These can be used to produce vaccines against, and in the
CC diagnosis of, tuberculosis (TB) infection. The present sequence is one of
CC the proteins of the invention.

| | | | |
|----|---|----------------|----|
| Oy | 1 | WNFAGIEAASAIOG | 15 |
| | | | |
| Db | 6 | WNFAGIEAASAIOG | 20 |

Search completed: July 3, 2003, 14:05:32
Job time : 31 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: July 3, 2003, 14:05:35 ; Search time 17.5 Seconds
(without alignments)
98.565 Million cell updates/sec

Title: US-09-830-839-1

Perfect score: 79

Sequence: 1 MTEQOWNFAGIEAAA 15

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 440863 seqs, 11492915 residues

Total number of hits satisfying chosen parameters: 440863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEM_PUB pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCr_NEM_PUB pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEM_PUB pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEM_PUB pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCrUS_PUBCOMB pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEM_PUB pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEM_PUB pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEM_PUB pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|------------------------|-------------------|
| 1 | 79 | 100.0 | 95 | 9 US-09-805-427A-1 | Sequence 1, Appli |
| 2 | 79 | 100.0 | 403 | 9 US-09-805-427A-4 | Sequence 4, Appli |
| 3 | 79 | 100.0 | 403 | 10 US-09-791-171-173 | Sequence 173, App |
| 4 | 79 | 100.0 | 404 | 9 US-09-805-427A-3 | Sequence 3, Appli |
| 5 | 79 | 100.0 | 404 | 10 US-09-791-171-172 | Sequence 172, App |
| 6 | 59 | 74.7 | 20 | 9 US-10-044-703-60 | Sequence 60, Appl |
| 7 | 59 | 74.7 | 20 | 10 US-09-813-333-59 | Sequence 59, Appl |
| 8 | 43 | 54.4 | 13 | 10 US-09-813-333-59 | Sequence 59, Appl |
| 9 | 43 | 54.4 | 13 | 10 US-09-813-333-59 | Sequence 59, Appl |
| 10 | 43 | 54.4 | 20 | 9 US-10-044-703-61 | Sequence 61, Appl |
| 11 | 43 | 54.4 | 20 | 10 US-09-813-333-61 | Sequence 61, Appl |
| 12 | 40 | 50.6 | 295 | 9 US-09-738-626-4782 | Sequence 4782, Ap |
| 13 | 39 | 49.4 | 916 | 10 US-09-815-242-12006 | Sequence 12006, A |
| 14 | 38 | 48.1 | 301 | 9 US-09-738-626-6880 | Sequence 6880, Ap |
| 15 | 38 | 48.1 | 410 | 10 US-09-847-0818-4 | Sequence 4, Appli |
| 16 | 38 | 48.1 | 438 | 9 US-10-198-070-38 | Sequence 38, Appl |
| 17 | 37 | 46.8 | 229 | 10 US-09-815-242-12391 | Sequence 12391, A |
| 18 | 37 | 46.8 | 325 | 9 US-09-738-626-6035 | Sequence 6035, Ap |
| 19 | 37 | 46.8 | 325 | 9 US-09-746-660A-4 | Sequence 4, Appli |

| | | | | | |
|----|------|------|------|------------------------|-------------------|
| 20 | 37 | 46.8 | 369 | 9 US-10-060-218A-6 | Sequence 6, Appli |
| 21 | 37 | 46.8 | 463 | 9 US-10-156-761-12653 | Sequence 12653, A |
| 22 | 37 | 46.8 | 739 | 9 US-10-156-761-10111 | Sequence 10111, A |
| 23 | 36 | 45.6 | 221 | 9 US-10-034-968-155 | Sequence 155, App |
| 24 | 36 | 45.6 | 221 | 10 US-09-739-254-155 | Sequence 155, App |
| 25 | 36 | 45.6 | 221 | 10 US-09-904-615-155 | Sequence 35, Appl |
| 26 | 36 | 45.6 | 511 | 10 US-09-833-745-35 | Sequence 54, Appl |
| 27 | 36 | 45.6 | 513 | 10 US-09-833-745-54 | Sequence 1148, A |
| 28 | 36 | 45.6 | 901 | 10 US-09-815-242-11148 | Sequence 1148, A |
| 29 | 36 | 45.6 | 3070 | 9 US-09-961-403-7 | Sequence 7, Appli |
| 30 | 35.5 | 44.9 | 593 | 10 US-09-920-021A-3 | Sequence 3, Appli |
| 31 | 35.5 | 44.9 | 597 | 9 US-10-038-010-22 | Sequence 22, Appl |
| 32 | 35 | 44.3 | 98 | 10 US-09-894-018-101 | Sequence 101, App |
| 33 | 35 | 44.3 | 153 | 10 US-09-925-301-1367 | Sequence 1367, Ap |
| 34 | 35 | 44.3 | 292 | 9 US-10-295-403-170 | Sequence 170, App |
| 35 | 35 | 44.3 | 301 | 9 US-10-028-072-176 | Sequence 176, App |
| 36 | 35 | 44.3 | 301 | 9 US-10-121-049-176 | Sequence 176, App |
| 37 | 35 | 44.3 | 301 | 9 US-10-123-904-176 | Sequence 176, App |
| 38 | 35 | 44.3 | 301 | 9 US-10-140-470-176 | Sequence 176, App |
| 39 | 35 | 44.3 | 301 | 9 US-10-175-746-176 | Sequence 176, App |
| 40 | 35 | 44.3 | 301 | 9 US-10-176-918-176 | Sequence 176, App |
| 41 | 35 | 44.3 | 301 | 9 US-10-176-921-176 | Sequence 176, App |
| 42 | 35 | 44.3 | 301 | 9 US-10-137-865-176 | Sequence 176, App |
| 43 | 35 | 44.3 | 301 | 9 US-10-140-474-176 | Sequence 176, App |
| 44 | 35 | 44.3 | 301 | 9 US-10-142-431-176 | Sequence 176, App |
| 45 | 35 | 44.3 | 301 | 9 US-10-143-114-176 | Sequence 176, App |

ALIGNMENTS

RESULT 1
US-09-805-427A-1
; Sequence 1, Application US/09805427A
; Patent No. US20020176867A1
; GENERAL INFORMATION:
; APPLICANT: Statens Serum Institut
; TITLE OF INVENTION: Hybrids of M. tuberculosis Antigens
; FILE REFERENCE: 670001-2002.5
; CURRENT APPLICATION NUMBER: US/09/805,427A
; CURRENT FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-805-427A-1

Query Match 100.0%; Score 79; DB 9; Length 95;
Best Local Similarity 100.0%; Pred. No. 3,4e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTEQOWNFAGIEAAA 15
|||||
DB 1 MTEQOWNFAGIEAAA 15
RESULT 2
US-09-805-427A-4
; Sequence 4, Application US/09805427A
; Patent No. US20020176867A1
; GENERAL INFORMATION:
; APPLICANT: Statens Serum Institut
; TITLE OF INVENTION: Hybrids of M. tuberculosis Antigens
; FILE REFERENCE: 670001-2002.5
; CURRENT APPLICATION NUMBER: US/09/805,427A
; CURRENT FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 403
; TYPE: PRT

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Fusion protein ESAT-6-Ag85B
US-09-805-427A-4
```

```
Query Match      100.0%; Score 79; DB 9; Length 403;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 MTEQOMNFAGIEAAA 15
        |||
Db      22 MTEQOMNFAGIEAAA 36
```

RESULT 3

```
US-09-791-171-173
; Sequence 173, Application US/09791171
; Patent No. US20020094336A1
; GENERAL INFORMATION:
; APPLICANT: ANDERSEN, Peter
; APPLICANT: NIELSEN, Rikke
; APPLICANT: OETTINGER, Thomas
; APPLICANT: RASMUSSEN, Peter Birk
; APPLICANT: ROSENKRANDS, Ida
; APPLICANT: WELDINGH, Karin
; TITLE OF INVENTION: NUCLEIC ACIDS FRAGMENTS AND POLYPEPTIDE FRAGMENTS
; FILE REFERENCE: 670001-2002.1
; CURRENT APPLICATION NUMBER: US/09/791,171
; CURRENT FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 09/050,739
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 0376/97
; PRIOR FILING DATE: 1997-04-02
; PRIOR APPLICATION NUMBER: 1277/97
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/044,624
; PRIOR FILING DATE: 1997-04-18
; PRIOR APPLICATION NUMBER: 60/070,488
; PRIOR FILING DATE: 1998-01-05
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 173
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-791-171-173
```

```
Query Match      100.0%; Score 79; DB 10; Length 403;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 MTEQOMNFAGIEAAA 15
        |||
Db      22 MTEQOMNFAGIEAAA 36
```

RESULT 4

```
US-09-805-427A-3
; Sequence 3, Application US/09805427A
; Patent No. US20020176867A1
; GENERAL INFORMATION:
; APPLICANT: Statens Serum Institut
; TITLE OF INVENTION: Hybrids of M. tuberculosis Antigens
; FILE REFERENCE: 670001-2002.5
; CURRENT APPLICATION NUMBER: US/09/805,427A
; CURRENT FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 404
; TYPE: PRT
```

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Fusion protein Ag85B-ESAT-6
US-09-805-427A-3
```

```
Query Match      100.0%; Score 79; DB 9; Length 404;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 MTEQOMNFAGIEAAA 15
        |||
Db      310 MTEQOMNFAGIEAAA 324
```

RESULT 5

```
US-09-791-171-172
; Sequence 172, Application US/09791171
; Patent No. US20020094336A1
; GENERAL INFORMATION:
; APPLICANT: ANDERSEN, Peter
; APPLICANT: NIELSEN, Rikke
; APPLICANT: OETTINGER, Thomas
; APPLICANT: RASMUSSEN, Peter Birk
; APPLICANT: ROSENKRANDS, Ida
; APPLICANT: WELDINGH, Karin
; TITLE OF INVENTION: NUCLEIC ACIDS FRAGMENTS AND POLYPEPTIDE FRAGMENTS
; FILE REFERENCE: 670001-2002.1
; CURRENT APPLICATION NUMBER: US/09/791,171
; CURRENT FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 09/050,739
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 0376/97
; PRIOR FILING DATE: 1997-04-02
; PRIOR APPLICATION NUMBER: 1277/97
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/044,624
; PRIOR FILING DATE: 1997-04-18
; PRIOR APPLICATION NUMBER: 60/070,488
; PRIOR FILING DATE: 1998-01-05
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 172
; LENGTH: 404
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-791-171-172
```

```
Query Match      100.0%; Score 79; DB 10; Length 404;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 MTEQOMNFAGIEAAA 15
        |||
Db      310 MTEQOMNFAGIEAAA 324
```

RESULT 6

```
US-10-044-703-60
; Sequence 60, Application US/10044703
; Publication No. US20020192233A1
; GENERAL INFORMATION:
; APPLICANT: Degroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/10/044,703
; CURRENT FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 60
```

; LENGTH: 20
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-044-703-60

Query Match 74.7%; Score 59; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 OWNFAGIEAAA 15
|||||
DB 1 OWNFAGIEAAA 11

RESULT 7
US-09-813-333-60
; Sequence 60, Application US/09813333
; Patent No. US20020119160A1
; GENERAL INFORMATION:

; APPLICANT: Degroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 1/999-004 US
; CURRENT APPLICATION NUMBER: US/09/813,333
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 60
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-813-333-60

Query Match 74.7%; Score 59; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 OWNFAGIEAAA 15
|||||
DB 1 OWNFAGIEAAA 11

RESULT 8
US-10-044-703-59
; Sequence 59, Application US/10044703
; Publication No. US20020192233A1
; GENERAL INFORMATION:

; APPLICANT: Degroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 1/999-004 US
; CURRENT APPLICATION NUMBER: US/10/044,703
; CURRENT FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 59
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-044-703-59

Query Match 54.4%; Score 43; DB 9; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 NFAGIEAAA 15
|||||
DB 1 NFAGIEAAA 9

RESULT 9

US-09-813-333-59
; Sequence 59, Application US/09813333
; Patent No. US20020119160A1
; GENERAL INFORMATION:

; APPLICANT: Degroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 1/999-004 US
; CURRENT APPLICATION NUMBER: US/09/813,333
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 59
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-813-333-59

Query Match 54.4%; Score 43; DB 10; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 NFAGIEAAA 15
|||||
DB 1 NFAGIEAAA 9

RESULT 10
US-10-044-703-61
; Sequence 61, Application US/10044703
; Publication No. US20020192233A1
; GENERAL INFORMATION:

; APPLICANT: Degroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 1/999-004 US
; CURRENT APPLICATION NUMBER: US/10/044,703
; CURRENT FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 61
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-044-703-61

Query Match 54.4%; Score 43; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.84;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 NFAGIEAAA 15
|||||
DB 1 NFAGIEAAA 9

RESULT 11
US-09-813-333-61
; Sequence 61, Application US/09813333
; Patent No. US20020119160A1
; GENERAL INFORMATION:

; APPLICANT: Degroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 1/999-004 US
; CURRENT APPLICATION NUMBER: US/09/813,333
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 61
; LENGTH: 20

;; TYPE: PRT
;; ORGANISM: Mycobacterium tuberculosis
US-09-813-333-61

Query Match 54.4%; Score 43; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.84;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 NFAGIEAAA 15
Db 1 NFAGIEAAA 9

RESULT 12
US-09-738-626-4782
; Sequence 4782, Application US/09738626
; Publication No. US20020197605A1

;; GENERAL INFORMATION:
;; APPLICANT: NAKAGAWA, SATOSHI
;; APPLICANT: MIZOGUCHI, HIROSHI
;; APPLICANT: ANDO, SEIKO
;; APPLICANT: HAYASHI, MIKIRO
;; APPLICANT: OCHIAI, KEIKO
;; APPLICANT: YOKOI, HARUHIKO
;; APPLICANT: TATEISHI, NAOKO
;; APPLICANT: SENOH, AKIHIRO
;; APPLICANT: IKEDA, MASATO
;; APPLICANT: OZAKI, AKIO
;; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;; FILE REFERENCE: 249-125
;; CURRENT APPLICATION NUMBER: US/09/738,626
;; CURRENT FILING DATE: 2000-12-18
;; PRIOR APPLICATION NUMBER: JP 99/377484
;; PRIOR FILING DATE: 1999-12-16
;; PRIOR APPLICATION NUMBER: JP 00/159162
;; PRIOR FILING DATE: 2000-04-07
;; PRIOR APPLICATION NUMBER: JP 00/280988
;; PRIOR FILING DATE: 2000-08-03
;; NUMBER OF SEQ ID NOS: 7059
;; SOFTWARE: PatentIn ver. 3.0
;; SEQ ID NO 4782
;; LENGTH: 295
;; TYPE: PRT
;; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4782

Query Match 50.6%; Score 40; DB 9; Length 295;
Best Local Similarity 55.6%; Pred. No. 40;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTEOQMNFA 9
Db 50 LTEEMNYA 58

RESULT 13
US-09-815-242-12006
; Sequence 12006, Application US/09815242
; Patent No. US20020061569A1
;; GENERAL INFORMATION:
;; APPLICANT: Haselbeck, Robert
;; APPLICANT: Ohlsen, Karl L.
;; APPLICANT: Zyskind, Judith W.
;; APPLICANT: Wall, Daniel
;; APPLICANT: Travick, John D.
;; APPLICANT: Carr, Grant J.
;; APPLICANT: Yamamoto, Robert T.
;; APPLICANT: Xu, H. Howard
;; TITLE OF INVENTION: Identification of Essential Genes in
;; FILE REFERENCE: ELITRA.011A
;; CURRENT APPLICATION NUMBER: US/09/815,242
;; CURRENT FILING DATE: 2001-03-21

;; PRIOR APPLICATION NUMBER: 60/191,078
;; PRIOR FILING DATE: 2000-03-21
;; PRIOR APPLICATION NUMBER: 60/206,848
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 12006
;; LENGTH: 916
;; TYPE: PRT
;; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-12006

Query Match 49.4%; Score 39; DB 10; Length 916;
Best Local Similarity 54.5%; Pred. No. 1.9e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 4 QOWNFAGIEAA 14
Db 699 EOWDIEGLEAA 709

RESULT 14
US-09-738-626-6880
; Sequence 6880, Application US/09738626
; Publication No. US20020197605A1
;; GENERAL INFORMATION:
;; APPLICANT: NAKAGAWA, SATOSHI
;; APPLICANT: MIZOGUCHI, HIROSHI
;; APPLICANT: ANDO, SEIKO
;; APPLICANT: HAYASHI, MIKIRO
;; APPLICANT: OCHIAI, KEIKO
;; APPLICANT: YOKOI, HARUHIKO
;; APPLICANT: TATEISHI, NAOKO
;; APPLICANT: SENOH, AKIHIRO
;; APPLICANT: IKEDA, MASATO
;; APPLICANT: OZAKI, AKIO
;; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;; FILE REFERENCE: 249-125
;; CURRENT APPLICATION NUMBER: US/09/738,626
;; CURRENT FILING DATE: 2000-12-18
;; PRIOR APPLICATION NUMBER: JP 99/377484
;; PRIOR FILING DATE: 1999-12-16
;; PRIOR APPLICATION NUMBER: JP 00/159162
;; PRIOR FILING DATE: 2000-04-07
;; PRIOR APPLICATION NUMBER: JP 00/280988
;; PRIOR FILING DATE: 2000-08-03
;; NUMBER OF SEQ ID NOS: 7059
;; SOFTWARE: PatentIn ver. 3.0
;; SEQ ID NO 6880
;; LENGTH: 301
;; TYPE: PRT
;; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6880

Query Match 48.1%; Score 38; DB 9; Length 301;
Best Local Similarity 55.6%; Pred. No. 90;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MTEOQMNFA 9
Db 57 LTEDENMYA 65

RESULT 15
 US-09-847-081B-4
 Sequence 4, Application US/09847081B
 Patent No. US20020128464A1
 GENERAL INFORMATION:
 APPLICANT: BAYER AG
 TITLE OF INVENTION: DNA encoding the tobacco phytoene synthase
 FILE REFERENCE: Lg A 34 326
 CURRENT APPLICATION NUMBER: US/09/847,081B
 CURRENT FILING DATE: 2001-05-02
 NUMBER OF SEQ ID NOS: 10
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 4
 LENGTH: 410
 TYPE: PRT
 ORGANISM: *Nicotiana tabacum*
 FEATURE:
 NAME/KEY: unsure
 LOCATION: 135..139
 OTHER INFORMATION: xaa is unknown or other
 US-09-847-081B-4

| | | | | | |
|----|-----------------------|-------|--------------------|--------|-------------------------|
| | Query Match | 48.1% | Score 38; | DB 10; | Length 410; |
| | Best Local Similarity | 41.7% | Pred. No. 1.2e+02; | | |
| | Matches | 5; | Conservative | 4; | Mismatches 3; Indels 0; |
| Qy | 3 EQOWNFACIEMA | 14 | | | |
| | : : : : | : | : | : | : |
| Db | 50 KÖRMNFGSVRSA | 61 | | | |

Search completed: July 3, 2003, 14:08:27
Job time : 18.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2003, 14:03:09 ; Search time 24 Seconds
(without alignments)
128.779 Million cell updates/sec

Title: US-09-830-839-1

Perfect score: 79

Sequence: 1 MTEQQWNPAGIEA 15

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_virus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 43 | 54.4 | 217 | 16 | Q9KMY8 |
| 2 | 43 | 54.4 | 247 | 16 | Q8X3F8 |
| 3 | 42 | 53.2 | 369 | 16 | Q9Z839 |
| 4 | 42 | 53.2 | 723 | 16 | Q9A279 |
| 5 | 41 | 51.9 | 182 | 2 | P95622 |
| 6 | 41 | 51.9 | 440 | 2 | Q9PAD1 |
| 7 | 41 | 51.9 | 443 | 2 | Q9AMK7 |
| 8 | 41 | 51.9 | 443 | 16 | Q9KVV6 |
| 9 | 41 | 51.9 | 471 | 17 | Q28786 |
| 10 | 41 | 51.9 | 934 | 16 | Q8XVJ6 |
| 11 | 40 | 50.6 | 38 | 16 | Q9KTC4 |
| 12 | 39 | 49.4 | 81 | 16 | Q98IF7 |
| 13 | 39 | 49.4 | 219 | 16 | Q8YVX4 |
| 14 | 39 | 49.4 | 282 | 2 | Q8K0L0 |
| 15 | 39 | 49.4 | 325 | 5 | Q966M0 |
| 16 | 39 | 49.4 | 328 | 16 | O51337 |

| | | | | | | |
|----|------|------|------|----|--------|---------------------|
| 17 | 39 | 49.4 | 393 | 16 | Q8YTF8 | Q8YTF8 anabaena sp |
| 18 | 39 | 49.4 | 482 | 10 | O04629 | O04629 arabidopsis |
| 19 | 39 | 49.4 | 598 | 10 | O8W4E7 | O8W4E7 lactococcus |
| 20 | 39 | 49.4 | 714 | 16 | Q9CE55 | Q9CE55 pseudomonas |
| 21 | 39 | 49.4 | 916 | 16 | Q9LCT3 | Q9LCT3 rhizobium 1 |
| 22 | 38.5 | 48.7 | 372 | 16 | Q98BQ8 | Q98BQ8 caulobacter |
| 23 | 38.5 | 48.7 | 392 | 16 | Q9A560 | Q9A560 caulobacter |
| 24 | 38 | 48.1 | 282 | 2 | Q9R6C8 | Q9R6C8 agrobacteri |
| 25 | 38 | 48.1 | 284 | 16 | Q9PH66 | Q9PH66 xyella fas |
| 26 | 38 | 48.1 | 319 | 16 | Q8UD90 | Q8UD90 agrobacteri |
| 27 | 38 | 48.1 | 328 | 10 | Q41401 | Q41401 sesbania ro |
| 28 | 38 | 48.1 | 388 | 2 | O52197 | O52197 azotobacter |
| 29 | 38 | 48.1 | 393 | 16 | Q9EWU0 | Q9EWU0 streptomyce |
| 30 | 38 | 48.1 | 415 | 16 | Q92005 | Q92005 rhizobium m |
| 31 | 38 | 48.1 | 443 | 16 | Q8ZN34 | Q8ZN34 salmonella |
| 32 | 38 | 48.1 | 443 | 16 | Q8Z4M1 | Q8Z4M1 salmonella |
| 33 | 38 | 48.1 | 806 | 3 | Q9P4T8 | Q9P4T8 agarticus bi |
| 34 | 38 | 48.1 | 885 | 2 | O46002 | O46002 clostridium |
| 35 | 38 | 48.1 | 2555 | 11 | Q91ZB5 | Q91ZB5 mus musculu |
| 36 | 38 | 48.1 | 2561 | 11 | Q8VIE5 | Q8VIE5 mus musculu |
| 37 | 37 | 46.8 | 61 | 16 | Q8X3Q2 | Q8X3Q2 escherichia |
| 38 | 37 | 46.8 | 108 | 17 | Q8TPE1 | Q8TPE1 methanosarc |
| 39 | 37 | 46.8 | 156 | 2 | Q93D85 | Q93D85 streptococc |
| 40 | 37 | 46.8 | 219 | 5 | Q95V33 | Q95V33 spodoptera |
| 41 | 37 | 46.8 | 229 | 16 | Q9SGS3 | Q9SGS3 staphylococ |
| 42 | 37 | 46.8 | 251 | 16 | Q9JYE0 | Q9JYE0 neisseria m |
| 43 | 37 | 46.8 | 251 | 16 | Q9JYB7 | Q9JYB7 neisseria m |
| 44 | 37 | 46.8 | 253 | 2 | O32816 | O32816 lactococcus |
| 45 | 37 | 46.8 | 256 | 16 | Q8YA41 | Q8YA41 listeria mo |

ALIGNMENTS

RESULT 1

Q9KMY8 PRELIMINARY; PRT; 217 AA.

ID Q9KMY8
AC Q9KMY8; 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-UN-2002 (TREMBLrel. 21, Last annotation update)
GN Frie protein.
DE Frie protein.
GN VCA0178.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=66;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L., Dodson R.D., Hatt D.H., Hickey E.K., Peterson J.D., Umayam L.A., Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D., Ermolaeva M.D., Vamathevan J., Baas S., Qin H., Dragoi I., Sellers P., McDonald L., Uitterback T., Fleischmann R.D., Nierman W.C., White O., Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C., Fraser C.M.;
RA "DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae."
RT Nature 406:477-483(2000).
DR EMBL: AE004358; AAF96091.1; -
DR TIGR: VCA0178; -
DR InterPro: IPR004287; HCCA_isomerase.
DR Pfam: PF03046; HCCA_isomerase; 1.
KW Complete proteome.
SQ SEQUENCE 217 AA; 24897 MW; 79B473E32313B5E8 CRC64;

Query Match 54.4%; Score 43; DB 16; Length 217;
Best Local Similarity 75.0%; Pred. No. 14;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 MTEQWNPAGIEA 13
||||| ||| |

DB 170 TEOQWLAGIHA 181

RESULT 2

0983F8 PRELIMINARY; PRT; 247 AA.

AC 08X3F8; MEDLINE-20330349; PubMed-10871362; Shiraishi M., Hiraoka H., Kimoto M., Tabuchi M., Kishi F., Ouchi K., Shiba T., Ishii K., Hattori M., Kihara S., Nakazawa T.; "Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWP029 from USA."; Nucleic Acids Res. 28:2311-2314(2000).

DT 01-MAR-2002 (TReMBLrel. 20, Created)

DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)

DE Hypothetical protein ECS1576.

GN ECS1576.

OS Escherichia coli O157:H7.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=83334;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / RIMD 0509952;

RA MEDLINE-21156231; PubMed-11258796; Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K., Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T., Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T., Kuhara S., Shiba T., Hattori M., Shinsagawa H.; Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12."; DNA Res. 8:11-22(2001).

RL DNA Res. 8:11-22(2001).

RT Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12.";

DR EMBL: AP002555; BAB34999.1; -.

DR InterPro: IPR000508; SigPase.

DR PROSITE: PS00761; SPASE_T_3; UNKNOWN_1.

KW Hypothetical protein.

SO SEQUENCE 247 AA; 27429 MW; 676680113259A870 CRC64;

QY 3 EQQWNPAGIEAA 15

DB 22 OQWNGFGGIKAA 34

Query Match 54.4%; Score 43; DB 16; Length 247;

Best Local Similarity 53.8%; Pred. No. 16;

Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

RESULT 3

092839 PRELIMINARY; PRT; 369 AA.

AC 092839; MEDLINE-10192388; PubMed-10192388; Lammell C., Fan J., Hyman R.W., Kallman S., Mitchell W., Marathe R., Stephens R.S.; "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."; Nat. Genet. 21:385-389(1999).

DT 01-MAY-1999 (TReMBLrel. 10, Created)

DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)

DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)

DE FE-8 oxidoreductase.

GN CPN0513 OR CPJ0513 OR CP0240.

OS Chlamydia pneumoniae (Chlamydia pneumoniae).

OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

OX NCBI_TaxID=83558;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CWL029;

RA MEDLINE-99206606; PubMed-10192388; Kallman S., Mitchell W., Marathe R., Lammell C., Fan J., Hyman R.W., Olinger L., Grimwood J., Davis R.W., Stephens R.S.; "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."; Nat. Genet. 21:385-389(1999).

RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";

RL Nat. Genet. 21:385-389(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=AR39;

RA MEDLINE-20150255; PubMed-10684935; Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F., White O., Hickey E.K., Peterson J., Uetliback T., Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L., Eisen J., Fraser C.M.; "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.";

RT pneumoniae AR39.";

RL Nucleic Acids Res. 28:1397-1406(2000).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=J138;

RA MEDLINE-20330349; PubMed-10871362; Shiraishi M., Hiraoka H., Kimoto M., Tabuchi M., Kishi F., Ouchi K., Shiba T., Ishii K., Hattori M., Kihara S., Nakazawa T.; "Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWP029 from USA."; Nucleic Acids Res. 28:2311-2314(2000).

DT 01-MAR-2002 (TReMBLrel. 17, Created)

DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)

DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)

DE Prolyl oligopeptidase family protein.

GN CC3687.

OS Caulobacter crescentus.

OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group; Caulobacter.

OX NCBI_TaxID=155892;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 19089 / CB15;

RA MEDLINE-21173698; PubMed-11259647; Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E., Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R., Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B., Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H., Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K., Uetliback T., Tran K., Wolf A., Yamathavan J., Ermolaeva M., White O., Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.; "Complete genome sequence of Caulobacter crescentus."; Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).

RL EMBL: AE006026; AAK25649.1; -.

DR TIGR: CC3687; -.

DR InterPro: IPR001375; Peptidase_S9.

DR InterPro: IPR004106; Peptidase_S9_N.

DR InterPro: IPR002470; Proligo_Phase.

DR InterPro: IPR000379; Ser esters_site.

DR Pfam: PF00326; Peptidase_S9_1.

DR Pfam: PF02897; Peptidase_S9_N; 1.

DR PRINTS: PR00862; PROLIGOPTASE.

KW Complete proteome.

SO SEQUENCE 723 AA; 78547 MW; 05F05150D74819A8 CRC64;

QY 2 TEOQWNPAGIEAA 15

DB 316 TEPWTFAGVAIPA 329

Query Match 53.2%; Score 42; DB 16; Length 723;

Best Local Similarity 50.0%; Pred. No. 72;

Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

RESULT 4

09A279 PRELIMINARY; PRT; 723 AA.

AC 09A279; MEDLINE-20330349; PubMed-10871362; Shiraishi M., Hiraoka H., Kimoto M., Tabuchi M., Kishi F., Ouchi K., Shiba T., Ishii K., Hattori M., Kihara S., Nakazawa T.; "Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWP029 from USA."; Nucleic Acids Res. 28:2311-2314(2000).

DT 01-MAR-2002 (TReMBLrel. 17, Created)

DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)

DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)

DE Prolyl oligopeptidase family protein.

GN CC3687.

OS Caulobacter crescentus.

OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group; Caulobacter.

OX NCBI_TaxID=155892;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 19089 / CB15;

RA MEDLINE-21173698; PubMed-11259647; Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E., Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R., Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B., Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H., Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K., Uetliback T., Tran K., Wolf A., Yamathavan J., Ermolaeva M., White O., Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.; "Complete genome sequence of Caulobacter crescentus."; Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).

RL EMBL: AE006026; AAK25649.1; -.

DR TIGR: CC3687; -.

DR InterPro: IPR001375; Peptidase_S9.

DR InterPro: IPR004106; Peptidase_S9_N.

DR InterPro: IPR002470; Proligo_Phase.

DR InterPro: IPR000379; Ser esters_site.

DR Pfam: PF00326; Peptidase_S9_1.

DR Pfam: PF02897; Peptidase_S9_N; 1.

DR PRINTS: PR00862; PROLIGOPTASE.

KW Complete proteome.

SO SEQUENCE 723 AA; 78547 MW; 05F05150D74819A8 CRC64;

QY 6 WNPAGIEAA 14

DB 305 WNYLGIEAA 313

Query Match 53.2%; Score 42; DB 16; Length 369;

Best Local Similarity 77.8%; Pred. No. 36;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

RESULT 5

Query Match 51.9%; Score 41; DB 2; Length 182;
Best Local Similarity 66.7%; Pred. No. 26;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 2 TEQOMNFRAG 10
51 TEEMTFAG 59

AC P95622 PRELIMINARY; PRT; 182 AA.
AC P95622;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Cytochrome c1 (Fragment).
OS Rhizobium leguminosarum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
CC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=384;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97158229; PubMed=9004501;
RA Wu G., Delgado M.J., Vargas C., Davies A.E., Poole R.K., Downie J.A.;
RT "The cytochrome bcl complex but not Cym is necessary for symbiotic
nitrogen fixation by Rhizobium leguminosarum.";
RL Microbiology 142:3381-3388(1996).
DR EMBL; X98018; CA66651.1; -;
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR002326; CytC.
DR Pfam; PF02167; Cytochrome_C1; 1.
DR PRINTS; PR00603; CYTOCHROME_C1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
FT NON_TER 182
SQ SEQUENCE 182 AA; 19787 MW; BC1C7255859D44B8 CRC64;

Query Match 51.9%; Score 41; DB 2; Length 182;
Best Local Similarity 66.7%; Pred. No. 26;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 2 TEQOMNFRAG 10
51 TEEMTFAG 59

AC P95622 PRELIMINARY; PRT; 440 AA.
AC P95622;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Cadr.
OS Cadr.
OC Moritella japonica.
OC Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;
CC Moritella.
OX NCBI_TaxID=89067;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=DSK1;
RA Li L., Fujii S., Kato C., Koki H.;
RT "Expression of Lysin Decarboxylase Gene Is Regulated by Pressure in a
RT Deep-Sea Piezophilic Bacterium, Moritella japonica.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB041735; BAB16717.1; -;
DR InterPro; IPR002293; AA/rel_pmeasel.
DR InterPro; IPR004841; Pemease.
DR Pfam; PF00324; aa_pemeases; 1.
SQ SEQUENCE 440 AA; 45577 MW; DBDCAADF8958CD9 CRC64;

Query Match 51.9%; Score 41; DB 2; Length 440;
Best Local Similarity 60.0%; Pred. No. 65;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 6 WNFAGIEAAA 15
199 WSFVGESAA 208

RESULT 7

Query Match 51.9%; Score 41; DB 2; Length 443;
Best Local Similarity 60.0%; Pred. No. 65;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 6 WNFAGIEAAA 15
199 WSFVGESAA 208

AC P95622 PRELIMINARY; PRT; 443 AA.
AC P95622;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Cadaverine/lysine antiporter.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RX Rhee J., Lee J., Lee S., Chung S., Rhee J., Choi S., Ryu P.;
RT "Vibrio vulnificus cadA genes.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF324470; AAK01663.1; -;
DR InterPro; IPR002293; AA/rel_pmeasel.
DR InterPro; IPR004841; Pemease.
DR Pfam; PF00324; aa_pemeases; 1.
SQ SEQUENCE 443 AA; 46752 MW; B88FD7E188036751 CRC64;

Query Match 51.9%; Score 41; DB 2; Length 443;
Best Local Similarity 60.0%; Pred. No. 65;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 6 WNFAGIEAAA 15
199 WSFVGESAA 208

AC P95622 PRELIMINARY; PRT; 445 AA.
AC P95622;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Cadaverine/lysine antiporter Cadr. putative.
GN VC0280.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragol I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Mierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
DR EMBL; AE004116; AAF93455.1; -;
DR TIGR; VC0280; -;
DR InterPro; IPR002293; AA/rel_pmeasel.
DR InterPro; IPR004841; Pemease.
DR Pfam; PF00324; aa_pemeases; 1.
KW Complete proteome.
SQ SEQUENCE 445 AA; 46907 MW; 3A64F21B9680684C CRC64;

RESULT 8

Query Match 51.9%; Score 41; DB 16; Length 445;
Best Local Similarity 60.0%; Pred. No. 65;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 6 WNFAGIEAAA 15
199 WSFVGESAA 208

AC P95622 PRELIMINARY; PRT; 445 AA.
AC P95622;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Cadaverine/lysine antiporter Cadr. putative.
GN VC0280.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragol I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Mierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
DR EMBL; AE004116; AAF93455.1; -;
DR TIGR; VC0280; -;
DR InterPro; IPR002293; AA/rel_pmeasel.
DR InterPro; IPR004841; Pemease.
DR Pfam; PF00324; aa_pemeases; 1.
KW Complete proteome.
SQ SEQUENCE 445 AA; 46907 MW; 3A64F21B9680684C CRC64;

DB 199 MSFVGESAA 208

RESULT 9

ID 028786 PRELIMINARY; PRT; 471 AA.

AC 028786;

DT 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Hypothetical protein AF1486.

GN AF1486.

OS Archaeoglobus fulgidus.

OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;

OC Archaeoglobaceae; Archaeoglobus.

OX NCBI_TaxID=2234;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;

RA MEDLINE=98049343; PubMed=9389475;

RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,

RA Ketchum K.A., Dodson R.J., Gwin M., Hickey E.K., Peterson J.D.,

RA Richardson D.L., Kurland J., Graham D.E., Kyrpides N.C.,

RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,

RA Kirschner E.F., Dougherty B.A., McKenney K., Adams M.D., Loftis B.,

RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,

RA Overbeek R., Cocayne J.D., Meldrum J.F., McDonald L., Uitterback T.,

RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,

RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,

RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,

RA Venter J.C.;

RT "The complete genome sequence of the hyperthermophilic, sulphate-

RT reducing archaeon Archaeoglobus fulgidus.";

RL Nature 390:364-370(1997).

DR EMBL: AEO01000; AAB89769.1; -

DR TIGR: AF1486; -

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 471 AA; 54165 MW; 892A03A47FBD4771 CRC64;

QY Query Match 51.9%; Score 41; DB 17; Length 471;

Best Local Similarity 61.5%; Pred. No. 69;

Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

DB 45 EQMWFEGKEADA 57

QY 3 EQQWNFAGIEA 15

DB 08XVJ6

RESULT 10

ID 08XVJ6 PRELIMINARY; PRT; 934 AA.

AC 08XVJ6;

DT 01-MAR-2002 (TREMBLrel. 20, Created)

DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Probable preprotein translocase SECA subunit.

GN SECA OR RSC2834 OR RS00271.

OS Ralstonia solanacearum (Pseudomonas solanacearum).

OC Ralstonia solanacearum (Pseudomonas solanacearum).

OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;

OC Ralstonia.

OX NCBI_TaxID=305;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=GM11000;

RA MEDLINE=21681879; PubMed=11823852;

RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,

RA Arlet M., Billault A., Brotlier P., Camus J.C., Catolico L.,

RA Chaudier M., Choisme N., Claudet-Renard C., Cunha S., Demange N.,

RA Gaspin C., Lawie M., Moisan A., Robert C., Saurin W., Schlex T.,

RA Siguler P., Thebaud P., Whalen M., Wincker P., Levy M.,

RA Weisenbach J., Boucher C.A.;

RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";

RL Nature 415:497-502(2002).

DR EMBL: AL646072; CAD16541.1; -

DR InterPro: IPR000185; SECA.

DR InterPro: IPR004027; SEC_C_motif.

DR Pfam: PF02810; SEC-C; 1.

DR Pfam: PF01043; SECA; protein; 1.

DR PRINTS: PR00906; SECA.

DR TIGRFAMs: TIGR00963; seca; 1.

DR PROSITE: PS01312; SECA; 1.

KW Complete proteome.

SQ SEQUENCE 934 AA; 105241 MW; F9C3FA601EA62095 CRC64;

QY Query Match 51.9%; Score 41; DB 16; Length 934;

Best Local Similarity 60.0%; Pred. No. 1,4e+02;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 708 EQQWDLAAGLE 717

QY 3 EQQWNFAGIE 12

RESULT 11

ID 09KTC4 PRELIMINARY; PRT; 38 AA.

AC 09KTC4;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE Hypothetical protein VC0978.

GN VC0978.

OS Vibrio cholerae.

OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.

OX NCBI_TaxID=666;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=EL TOR N16961 / SEROTYPE O1;

RA MEDLINE=20406833; PubMed=10952301;

RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,

RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,

RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,

RA Emolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,

RA McDonald L., Uitterback T., Fleischmann R.D., Nierman W.C., White O.,

RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,

RA Fraser C.M.;

RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio

RT cholerae.";

RL Nature 406:477-483(2000).

DR EMBL: AEO04179; AAF94140.1; -

DR TIGR: VC0978; -

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 38 AA; 4427 MW; 3077749E7004F910 CRC64;

QY Query Match 50.6%; Score 40; DB 16; Length 38;

Best Local Similarity 50.0%; Pred. No. 7.9;

Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

DB 18 MTLNQMWIGVGA 31

QY 1 MTEQWNFAGIEA 14

DB 098IF7

RESULT 12

ID 098IF7 PRELIMINARY; PRT; 81 AA.

AC 098IF7;

DT 01-OCT-2001 (TREMBLrel. 18, Created)

DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Hypothetical protein msr2423.

GN MSR2423.

OS Rhizobium loti (Mesorhizobium loti).

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Phyllobacteriaceae; Mesorhizobium.

OX NCBI_TaxID=381;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2003, 14:00:24 ; Search time 30 Seconds
(Without alignments)
66.625 Million cell updates/sec

Title: US-09-830-839-1
Perfect score: 79
Sequence: 1 MTEQOMNFAGIEAAA 15

Scoring table:
BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: /_Geneseq_101002: *
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT: *
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT: *
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT: *
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT: *
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT: *
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT: *
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT: *
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT: *
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT: *
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT: *
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT: *
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT: *
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT: *
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT: *
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT: *
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT: *
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT: *
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT: *
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT: *
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT: *
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT: *
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT: *
24: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length DB | ID | Description |
|------------|-------|-------------|-----------|----------|--------------------|
| 1 | 79 | 100.0 | 15 | AA194590 | Mycobacterium tube |
| 2 | 79 | 100.0 | 15 | AA194590 | Early secreted ant |
| 3 | 79 | 100.0 | 17 | AA194590 | Antigenic N-termin |
| 4 | 79 | 100.0 | 19 | AA194590 | ESAT-6 antigen SEQ |
| 5 | 79 | 100.0 | 51 | AA194590 | Mycobacterium tube |
| 6 | 79 | 100.0 | 51 | AA194590 | Mycobacterium tube |
| 7 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 8 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 9 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 10 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |

| | | | | | |
|----|----|-------|----|----------|--------------------|
| 11 | 79 | 100.0 | 51 | AA194590 | Mycobacterium tube |
| 12 | 79 | 100.0 | 51 | AA194590 | Early secreted ant |
| 13 | 79 | 100.0 | 51 | AA194590 | Antigenic N-termin |
| 14 | 79 | 100.0 | 51 | AA194590 | ESAT-6 antigen SEQ |
| 15 | 79 | 100.0 | 51 | AA194590 | Mycobacterium tube |
| 16 | 79 | 100.0 | 51 | AA194590 | Mycobacterium tube |
| 17 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 18 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 19 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 20 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 21 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 22 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 23 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 24 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 25 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 26 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 27 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 28 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 29 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 30 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 31 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 32 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 33 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 34 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 35 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 36 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 37 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 38 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 39 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 40 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 41 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 42 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 43 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 44 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |
| 45 | 79 | 100.0 | 51 | AA194590 | M. tuberculosis |

ALIGNMENTS

| | |
|----------|---|
| RESULT 1 | |
| AA194590 | standard; peptide: 15 AA. |
| ID | AA194590 |
| XX | AA194590 |
| AC | AA194590 |
| DT | 28-NOV-2000 (first entry) |
| XX | Mycobacterium tuberculosis ESAT-6 peptide ES1. |
| DE | Mycobacterium tuberculosis ESAT-6 peptide ES1. |
| XX | Tuberculosis; infection diagnosis; ESAT-6. |
| KW | Tuberculosis; infection diagnosis; ESAT-6. |
| OS | Mycobacterium tuberculosis. |
| XX | MO200026248-A2. |
| PN | 11-MAY-2000. |
| XX | 03-NOV-1999; 99WO-GB03635. |
| PF | 04-NOV-1998; 98GB-0024213. |
| XX | 04-NOV-1998; 98US-0107004. |
| PR | (ISIS-) ISIS INNOVATION LTD. |
| PA | Lalvani A, Pathan AA; |
| XX | WPI: 2000-365579/31. |
| DR | Novel method of diagnosing infection, or exposure of a host, to a |
| XX | mycobacterium comprising contacting T cells from the host with ESAT-6 |
| PT | derived peptides - |
| XX | |

Mycobacterium tube
Mycobacterium spec
ESAT-6. Mycobacter
Mycobacterium tube
Mycobacterium tube
M. tuberculosis RV3
Mycobacterium tube
Mycobacterium tube
Mycobacterium tube
Mycobacterium tube
Amino acid sequenc
Mycobacterium tube
Mycobacterium tube
Early secreted ant
N-terminal amino a
N-terminal peptide
Mycobacterium tube
Mycobacterium tube
Chlamydia pneumoni
Phytoene synthase
Amino acid sequenc
C glutamicum prote
Corynebacterium g1
Novel human diagno
Propionibacterium
Arabidopsis thalia
Arabidopsis thalia
Amino acid sequenc
Human EST encoded
Arabidopsis thalia
Arabidopsis thalia
Herbically activ
Streptococcus poly
Streptococcus poly

PS Claim 1; Page 3; 33pp; English.

XX
CC The present sequence is the peptide ES1 derived from the Mycobacterium
CC tuberculosis ESAT-6 gene. This sequence is one of eleven peptides
CC derived from the ESAT-6 gene (see AAY94590 to AAY94600). The peptides
CC are recognised, to varying degrees, by the T cells of patients with
CC tuberculosis. When the peptides are combined together as a panel they
CC provide a high specificity and sensitivity diagnostic test for
CC M. tuberculosis infection. This test has the advantage that it does not
CC give false positives for patients who have been vaccinated with BCG.

XX Sequence 15 AA;

Query Match 100.0%; Score 79; DB 21; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.2e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MTEOQWNPAGIEAAA 15

DB 1 MTEOQWNPAGIEAAA 15

RESULT 2

ID AAO17432 standard; peptide; 15 AA.

XX AAO17432;

DT 11-JUL-2002 (first entry)

DE Early secreted antigenic target 6 T cell epitope #1.

XX Tuberculosis; TB; vaccination; vaccine; CD4+ T cell immune response;

KW poxvirus vector; HIV; malaria; Helicobacter pylori; influenza;

KW hepatitis; viral infection; leprosy; protozoan parasite; cancer;

KW tuberculosis; anti-HIV; protozoicide; antibacterial; virucide;

XX hepatotropic; antiinflammatory; antileprotic; cytostatic; epitope.

XX Mycobacterium tuberculosis.

XX WO200224224-A2.

XX 28-MAR-2002.

XX 13-SEP-2001; 2001WO-GB04116.

XX 21-SEP-2000; 2000GB-0023203.

XX (OXO-) OXON PHARMACINES LTD.

XX HILL AVS, McShane H, Gilbert S, Reece W, Schneider J;

XX WPI; 2002-394098/42.

XX Inducing CD4+ T-cell response against target antigen by administering a

XX composition comprising a source of CD4+ epitopes which is a

XX non-replicating or replication impaired recombinant poxvirus vector

XX Example 1; Page 21; 50pp; English.

XX The present invention relates to a method of inducing a CD4+ T-cell
XX response against a target antigen, by administering two different
XX compositions comprising a source of CD4+ T-cell epitope(s) of the target
XX antigen, where the second composition further includes an epitope which
XX is same as the epitope of the first composition, where the source of the
XX epitopes for the compositions is a non-replicating or replication
XX impaired recombinant poxvirus vector. The methods are useful for inducing
XX CD4+ T-cell immune responses against diseases such as tuberculosis, human
XX immunodeficiency virus (HIV) (persistent viral infection), malaria,
XX Helicobacter pylori, influenza, hepatitis (chronic hepatitis B and C),
XX cytomegalovirus (CMV), viral infection, herpes virus-induced disease,
XX leprosy and diseases caused by non-malarial protozoan parasite such as
XX toxoplasma and cancer. The present sequence is an epitope from

CC Mycobacterium tuberculosis useful in the method of the invention.

XX Sequence 15 AA;

Query Match 100.0%; Score 79; DB 23; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.2e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MTEOQWNPAGIEAAA 15

DB 1 MTEOQWNPAGIEAAA 15

RESULT 3

ID AAY88581 standard; peptide; 17 AA.

XX AAY88581;

DT 11-AUG-2000 (first entry)

DE Antigenic N-terminal sequence of ESAT-6.

XX Ligand presenting assembly; early secreted antigen target 6; ESAT-6;

KW bacteria; detect; diagnosis; allergen; cancer; vaccine; immune response;

KW neurotropic factor; autoimmune-system related compound; IPA; fungi;

XX parasite; cell-adhesion molecule.

XX Mycobacterium tuberculosis.

XX Key Location/Qualifiers

FT Cross-links 17

FT /note- "Ala at position 17 is linked via a linker to Ala

FT at position 17 of an identical peptide, where

FT the linker is COCH2-CH(NH-lys)-CH2CO, and the

FT Lys residue contained in the linker is also

FT linked to another peptide (see AAY8579)."
 PN WO200018791-A1.

XX 06-APR-2000.

XX 29-SEP-1999; 99WO-DK00510.

XX 29-SEP-1998; 98DK-0001233.

XX (STAT-) STAPENS SERUM INST.

XX (HOLM/) HOLM A.

XX Holm A, Jorgensen RM, Ostergaard S, Theisen M;

XX WPI; 2000-303438/26.

XX New ligand presenting assemblies useful for diagnosis, treatment and

XX prevention of diseases caused by e.g. viruses, bacteria, toxins,

XX PT allergens, autoimmune system-related compounds, cancer-related

XX compounds, cell adhesion molecules

XX Claim 34; Page 81; 100pp; English.

XX This sequence represents an antigenic N-terminal sequence from the
XX Mycobacterium tuberculosis early secreted antigenic target 6 (ESAT-6)
XX protein. The peptide is presented on the new ligand presenting assembly
XX of the invention. The invention relates to method for preparing ligand
XX presenting assemblies (LPAs) comprising:
XX (a) providing by solid phase synthesis, or fragment coupling, ligands
XX comprising desired sequences (e.g. the present sequence), the ligands
XX being attached to a solid phase;
XX (b) if necessary, deprotecting any N-terminal amino groups while the
XX ligands are still attached to the solid phase;
XX (c) reacting the ligands having unprotected N-terminal amino groups with
XX an achiral di-, tri- or tetracarboxylic acid, to provide a construct
XX having a ring structure; and

(d) cleaving the construct from the solid phase, to provide an LPA comprising ligands having free C-terminal groups. The LPAs can be used for raising an immune response in an animal. They can also be used in vaccines and for generating antibodies in an animal. Alternatively they can be used for the treatment, alleviation, detection, diagnosis, or prophylaxis of diseases caused by viruses, bacteria, toxins, allergens, autoimmune system-related compounds, cancer related compounds, cell adhesion molecules, neurotropic factors, fungi or parasites. Use of the method enables the preparation of very long ring systems interconnected by reaction with the achiral di-, tri- or tetracarboxylic acid. The ring structure formed between desired sequences further enables additional presentation of desired sequences and chemical moieties. The LPAs provide very flexible systems for polyfunctional constructs, and furthermore, products of high purity are obtained.

Sequence 17 AA;

Query Match 100.0%; Score 79; DB 21; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTEQOWNFAGIEAAA 15
|||||
DB 1 MTEQOWNFAGIEAAA 15

RESULT 4

AAW35548
ID AAW35548 standard; peptide; 19 AA.

AC AAW35548;

DT 22-APR-1998 (first entry)

DE ESAT-6 antigen SEQ ID NO:120 from WO9738011.

OS T-cell stimulatory peptide; immunogen; non-dendritic; carrier; tumour; scaffold; inhibition; metastasis; wound healing; solid phase.

PN WO9738011-A1.

PD 16-OCT-1997.

PF 03-APR-1997; 97WO-DE00146.

PR 03-APR-1996; 96DK-0000398.

PA (PEPR-) PEPRESEARCH AS.

PI Heegaard PMH, Jakobsen PH;

DR WPI; 1997-512645/47.

XX Non-dendritic peptide carrier linked to a solid phase - useful as a

PT diagnostic agent and as a scaffold for production of chemical

PT derivatives

PS Example 32; Page 156; 262pp; English.

CC A non-dendritic peptide carrier (A) has been developed which is coupled through a linker to a solid phase, forming a complex of (A)-solid phase. Where (A) comprises 10-50 amino acids capable of forming a secondary structure in a benign buffer after liberation from the solid phase, and further the (A)-solid phase complex comprises an immunogenic substance and/or an immune mediator coupled on (A). The present sequence CC represents a peptide used in an example from the present invention. An (A)-solid phase complex can be used as a scaffold for the production of chemical derivatives, characterised by covalently attaching molecules at attachment points. Alternatively (A) is used as a scaffold-peptide for the incorporation into an immunostimulating complex (iscom) resulting an (A)-iscom complex which is used for the chemical coupling of antigenic

CC substances in an aqueous solution by conjugation. (A) derivatised with one or more peptides having fibronectin-, laminin- or vitronectin-like binding activities can be used for the promotion of cell-attachment to plastic surfaces, in particular to inhibit tumour growth and metastasis, and for promotion of wound healing. Also a derivatised (A) can be used for the selection of specifically-binding aptamers or as a diagnostic agent. Such diagnostic-(A) molecules could be used to detect molecules derived from or indicative of pregnancy or of a disease, such as an infectious, autoimmune or cancerous disease.

Sequence 19 AA;

Query Match 100.0%; Score 79; DB 18; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.5e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTEQOWNFAGIEAAA 15
|||||
DB 1 MTEQOWNFAGIEAAA 15

RESULT 5

AAW32466
ID AAW32466 standard; Protein; 51 AA.

AC AAW32466;

DT 09-JAN-1998 (first entry)

DE Mycobacterium tuberculosis antigen ESAT-6.

OS Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;

PN WO9709428-A2.

PD 13-MAR-1997.

PF 30-AUG-1996; 96WO-US14674.

PR 12-JUL-1996; 96US-0680574.

PR 01-SEP-1995; 95US-0523436.

PR 22-SEP-1995; 95US-0533634.

PR 22-MAR-1996; 96US-0620874.

PR 05-JUN-1996; 96US-0659683.

PA (CORI-) CORIXA CORP.

PI Campos-melo A, Dillon DC, Houghton R, Reed SG, Skelky YAW;

DR Twardzik DR, Vedvick TH;

DR WPI; 1997-192903/17.

XX N-PSDB; AAT91529.

PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are

PT useful in vaccines for prevention or treatment of tuberculosis, also

PT for diagnosis

PS Disclosure; Page 135; 168pp; English.

CC A new immunogenic polypeptide has been developed comprising an immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its variant differing only in conservative substitutions and/or modifications). The present sequence represents a M.tuberculosis antigen, ESAT-6. The immunogenic protein, and fusion proteins containing one or more of the proteins or one of the proteins plus a ESAT-6, are useful in vaccines, preferably when formulated with a non-specific adjuvant, to induce an immune response against M.tuberculosis (for treatment or prevention).

Query Match 100.0%; Score 79; DB 18; Length 51;
 Best Local Similarity 100.0%; Pred. No. 4.2e-06;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MTEOQMNFGIEFAAA 15
 |||||
 DB 1 MTEOQMNFGIEFAAA 15

RESULT 6

AAW32339
 ID AAW32339 standard; Protein; 51 AA.

AC AAW32339;

DT 13-JAN-1998 (first entry)

DE Mycobacterium tuberculosis antigen ESAT-6.

KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
 KM skin testing; M. tuberculosis.

OS Mycobacterium tuberculosis.

PN WO9709429-A2.

PD 13-MAR-1997.

PF 30-AUG-1996; 96WO-US14675.

PR 12-JUL-1996; 96US-0680573.

PR 01-SEP-1995; 95US-0523435.

PR 22-SEP-1995; 95US-0532136.

PR 22-MAR-1996; 96US-0620280.

PR 05-JUN-1996; 96US-0658800.

PS (CORI-) CORIXA CORP.

PI Campos-neto A, Dillon DC, Houghton R, Reed SC, Skelky YAW,
 Twardzik DR, Vedvick TH;

DR WPI: 1997-192904/17.

DR N-PSDB: AAT91463.

PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens

PT - useful for diagnosis of M. tuberculosis infection

PS Claim 43; Page 147; 190pp; English.

CC A new immunogenic polypeptide has been developed comprising an

CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or

CC its variant differing only in conservative substitutions and/or

CC modifications). The present sequence represents a specifically claimed

CC sequence of the ESAT-6 M. tuberculosis antigen. The immunogenic

CC polypeptide can be used to diagnose M. tuberculosis infection by forming

CC complexes with specific antibodies in the sample. Fragments of DNA

CC encoding the immunogenic polypeptide can be used as diagnostic primers

CC or probes and agents that bind to the antigen, especially monoclonal

CC antibodies or equivalent polyclonal antibodies, are also used for

CC diagnosis.

CC Sequence 51 AA;

Query Match 100.0%; Score 79; DB 18; Length 51;

Best Local Similarity 100.0%; Pred. No. 4.2e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MTEOQMNFGIEFAAA 15
 |||||
 DB 1 MTEOQMNFGIEFAAA 15

RESULT 7
 AAW81701
 ID AAW81701 standard; Protein; 51 AA.

AC AAW81701;

DT 27-JAN-1999 (first entry)

DE M. tuberculosis immunogenic polypeptide ESAT-6.

KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;

KW vaccine; pharmaceutical; infection; diagnosis.

OS Mycobacterium tuberculosis.

PN WO9816646-A2.

PD 23-APR-1998.

PF 07-OCT-1997; 97WO-US18293.

PR 13-MAR-1997; 97US-0818112.

PR 11-OCT-1996; 96US-0730510.

PS (CORI-) CORIXA CORP.

PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
 Reed SC, Skelky YAW, Twardzik DR, Vedvick TS;

DR WPI: 1998-261042/23.

DR N-PSDB: AAV64501.

PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used

PT to develop products for the detection of M. tuberculosis infection

PT and for diagnosis, treatment and prevention of tuberculosis

PS Disclosure; Page 126; 230pp; English.

CC This sequence represents an immunogenic portion of a soluble

CC Mycobacterium tuberculosis (MT) antigen which can be used in a method

CC for inducing protective immunity against tuberculosis (TB). This

CC sequence can be formulated into vaccines and/or pharmaceutical

CC compositions for immunising against M. tuberculosis infection or may

CC be used for the diagnosis of tuberculosis.

CC Sequence 51 AA;

Query Match 100.0%; Score 79; DB 19; Length 51;

Best Local Similarity 100.0%; Pred. No. 4.2e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MTEOQMNFGIEFAAA 15
 |||||
 DB 1 MTEOQMNFGIEFAAA 15

RESULT 8

AAW64334
 ID AAW64334 standard; Protein; 51 AA.

AC AAW64334;

DT 09-NOV-1998 (first entry)

DE Mycobacterium tuberculosis antigen ESAT-6.

KW Tuberculosis; infection; diagnosis; antigen; ESAT-6.

OS Mycobacterium tuberculosis.

PN WO9816645-A2.

PD 23-APR-1998.

XX 07-OCT-1997; 97WO-US18214.
XX 13-MAR-1997; 97US-0818111.
PR 11-OCT-1996; 96US-0729622.
XX
PA (CORI-) CORIXA CORP.
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;
XX
DR WPI: 1998-251292/22.
DR N-PSDB; AA44393.
XX
PT New isolated Mycobacterium tuberculosis polyptides and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and diagnosis of tuberculosis
XX
PS Disclosure; Page 130; 250pp; English.
XX
CC This polypeptide comprises an antigenic portion of Mycobacterium
CC tuberculosis antigen ESAT-16. The invention relates to
CC compositions and methods for diagnosing tuberculosis. It provides
CC polypeptides (see AA64291-W64379) comprising an antigenic portion
CC of a soluble M. tuberculosis antigen, or an immunogenic portion of
CC an M. tuberculosis antigen, as well as fusion proteins between
CC these polypeptides and known antigens such as ESAT-6. Also claimed
CC are methods and diagnostic kits for detecting M. tuberculosis
CC infection in a patient using these polypeptides, antibodies, or
CC oligonucleotide probes and primers.
XX
SQ Sequence 51 AA;

Query Match 100.0%; Score 79; DB 19; Length 51;
Best Local Similarity 100.0%; Pred. No. 4.2e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MTEQOMNFAGIEAAA 15
DB 1 MTEQOMNFAGIEAAA 15

RESULT 9
AA39131
ID AA39131 standard; Protein; 51 AA.
XX
AC AA39131;
XX
DT 05-NOV-1999 (first entry)
XX
DE M. tuberculosis ESAT-6 amino acid sequence.
XX
KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9942076-A2.
XX
PD 26-AUG-1999.
XX
PE 17-FEB-1999; 99WO-US03268.
XX
PR 05-MAY-1998; 98US-0072967.
PR 18-FEB-1998; 98US-0025197.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;
XX
DR WPI: 1999-527409/44.

DR N-PSDB; AA219303.
XX
XX New antigens from Mycobacterium tuberculosis useful in diagnostic
PT skin tests and protective or therapeutic vaccines or compositions
XX
PS Disclosure; Page 120; 299pp; English.
XX
CC The present invention describes polypeptides comprising an immunogenic
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
CC are vaccines and fusion protein containing M. tuberculosis Ag's.
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
CC other polypeptides fragments, can be used in pharmaceutical compositions
CC or vaccines to generate a protective or therapeutic immune response to
CC M. tuberculosis and as reagents in skin tests for diagnosis of
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
CC by, T, B or natural killer cells and/or macrophages in
CC tuberculosis-immune subjects. AA219249 to AA219460 and AA39083 to
CC AA39225 are used in the exemplification of the present invention.
XX
SQ Sequence 51 AA;

Query Match 100.0%; Score 79; DB 20; Length 51;
Best Local Similarity 100.0%; Pred. No. 4.2e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MTEQOMNFAGIEAAA 15
DB 1 MTEQOMNFAGIEAAA 15

RESULT 10
AA38988
ID AA38988 standard; Protein; 51 AA.
XX
AC AA38988;
XX
DT 05-NOV-1999 (first entry)
XX
DE M. tuberculosis recombinant antigen protein ESAT-6.
XX
KW Antigen; diagnosis; detection; infection; antibody; immunisation;
KW vaccine; immunity.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9942118-A2.
XX
PD 26-AUG-1999.
XX
PE 17-FEB-1999; 99WO-US03265.
XX
PR 05-MAY-1998; 98US-0072596.
PR 18-FEB-1998; 98US-0024753.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;
XX
DR WPI: 1999-527416/44.
DR N-PSDB; AA219091.
XX
PT New polypeptide comprising antigenic portions of M. tuberculosis
XX
PS Claim 51; Page 165; 323pp; English.
XX
CC This invention describes novel recombinant antigens and their encoding
CC nucleic acids derived from Mycobacterium tuberculosis. The novel
CC polypeptides are useful for detecting M. tuberculosis infection in a
CC biological sample by detecting antibodies which bind with the
CC polypeptides, and are useful as vaccines for immunizing against
CC M. tuberculosis infection. The new detection methods are needed as
CC current vaccination strategies do not provide 100% immunity.

| Query Match | 100.0%; | Score 79; | DB 20; | Length 51; |
|--------------------------|--|---------------------------|-----------|------------|
| Best Local Similarity | 100.0%; | Pred. No. 4.2e-06; | | |
| Matches 15; Conservative | 0; | Mismatches 0; | Indels 0; | Gaps 0; |
| OY | 1 MTEOQWNPAGIEAAA 15 | | | |
| DB | 1 MTEOQWNPAGIEAAA 15 | | | |
| RESULT 11 | | | | |
| ID | AAU01897 | standard; Protein; 51 AA. | | |
| AC | AAU01897; | | | |
| DX | 29-AUG-2001 (first entry) | | | |
| DE | Mycobacterium tuberculosis partial antigen ESAT-6. | | | |
| XX | ESAT-6; antigen; vaccine; tuberculosis; AIDS; acquired immunodeficiency disease. | | | |
| XX | Mycobacterium tuberculosis. | | | |
| XX | WO200124820-A1. | | | |
| XX | 12-APR-2001. | | | |
| XX | 10-OCT-2000; 2000WO-US28095. | | | |
| XX | 07-OCT-1999; 99US-0158338. | | | |
| XX | 07-OCT-1999; 99US-0158425. | | | |
| XX | (CORI-) CORIXA CORP. | | | |
| XX | Skelky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes MT. | | | |
| XX | WPI; 2001-290576/30. | | | |
| XX | N-PSDB; AAS03787. | | | |
| XX | Vaccinating against Mycobacteria infections in mammals using fusion proteins comprising combinations of heterologous antigens | | | |
| XX | Disclosure; Page 164; 168pp; English. | | | |
| XX | The sequence represents Mycobacterium tuberculosis ESAT-6 antigen. Compositions comprising at least 2 heterologous antigens, as a fusion protein, and vectors expressing the fusion proteins are used as vaccines to prophylactically immunise mammals (especially humans) against infection by Mycobacteria. The compositions contain at least 2 heterologous antigens that increase the serological sensitivity of individuals infected with tuberculosis, a disease frequently affecting patients with acquired immunodeficiency disease, AIDS. | | | |
| XX | Sequence 51 AA: | | | |
| OY | 1 MTEOQWNPAGIEAAA 15 | | | |
| DB | 1 MTEOQWNPAGIEAAA 15 | | | |
| RESULT 12 | | | | |
| ID | AAE17580 | standard; Protein; 51 AA. | | |
| XX | AAE17580; | | | |

[illegible]

DE ESAT6.
 XX
 KW ESAT6: Mycobacterium tuberculosis: 6kDa antigen; tuberculosis complex;
 KW HYB76-8 reactive antigen; interferon-gamma; memory T-lymphocyte; vaccine;
 KW M. africanum; M. bovis; delayed-type hypersensitivity reaction; therapy.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN MO9501441-A1.
 XX
 PD 12-JAN-1995.
 XX
 PF 01-JUL-1994; 94MO-DK00273.
 XX
 PR 02-JUL-1993; 93DK-0000798.
 XX
 PA (STAT-) STATENS SERUMINSTITUTT.
 XX
 PI Andersen AB, Andersen P, Haslov K, Sorensen A;
 XX
 DR WPI: 1995-061005/08.
 DR N-PSDB: AAT51422.
 XX
 XX Vaccine for tuberculosis induces interferon-gamma release from
 PT T-lymphocytes - comprises an antigen released from mycobacteria,
 PT for immunisation of humans
 XX
 PS Claim 5; Page 61-63; 101pp; English.
 CC
 CC This sequence represents the Mycobacterium tuberculosis ESAT6. ESAT6 is
 CC also known as the 6kDa antigen, or the HYB76-8 reactive antigen. ESAT6
 CC is released from metabolising bacteria, and can be isolated from short
 CC term filtrates grown as shaken cultures for 7 days. ESAT6 also induces a
 CC release of interferon-gamma from reactivated memory T-lymphocytes. This
 CC protein sequence is included in the vaccine of the invention. The
 CC vaccine is for immunising an animal (including humans) against
 CC tuberculosis caused by a Mycobacteria belonging to the tuberculosis
 CC complex. The Mycobacteria of the tuberculosis complex are, M.
 CC tuberculosis, M. africanum, and M. bovis. The vaccine can evoke a
 CC protective immune response against tuberculosis or a delayed-type
 CC hypersensitivity reaction. The protein can also be included in a
 CC composition for diagnosing tuberculosis. The composition is injected
 CC intradermally, and a skin reaction is an indicator of tuberculosis.
 CC
 SQ Sequence 95 AA;
 QY
 Db 1 MTEOQWNPAGIEAANA 15
 |||||
 1 MTEOQWNPAGIEAANA 15
 RESULT 14
 AAY29890
 ID AAY29890 standard; Protein: 95 AA.
 XX
 AC AAY29890;
 XX
 DT 18-NOV-1999 (first entry)
 XX
 DE Mycobacterium tuberculosis ESAT-6 protein sequence.
 XX
 KW Mycobacterial; lactic acid bacterium; diagnosis; skin test; vaccine;
 KW delayed type hypersensitivity; DTB; ESAT-6 homodimer; tuberculosis;
 KW interferon-gamma release.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN MO9945119-A2.
 XX

PD 10-SEP-1999.
 XX
 PF 05-MAR-1999; 99MO-DK00109.
 XX
 PR 06-MAR-1998; 98DK-0000306.
 PR 06-MAR-1998; 98US-0077105.
 XX
 PA (STAT-) STATENS SERUM INST.
 XX
 PI Jensen CL, Folkersen J;
 XX
 DR WPI: 1999-551043/46.
 DR N-PSDB: AAZ21132.
 XX
 XX New mycobacterial polypeptide produced in lactic acid bacteria, useful
 PT in tuberculosis diagnosis and vaccines
 XX
 PS Disclosure; Page 75-76; 76pp; English.
 CC
 CC The present invention describes a bioactive polypeptide (or
 CC immunologically equivalent analogue) produced in lactic acid bacteria
 CC which reacts with lymphoid cells primed with Mycobacterium tuberculosis
 CC complex mycobacteria (M. tuberculosis, M. africanum or M. bovis). The
 CC polypeptide and ESAT-6 polypeptides are useful in compositions for
 CC diagnosis of and vaccination against tuberculosis caused by
 CC M. tuberculosis complex mycobacteria. The ESAT-6 polypeptide can be used
 CC to diagnose ongoing/previous sensitisation with these bacteria by
 CC detecting cytokine release when contacting blood samples with the
 CC polypeptide. The bioactive polypeptide may be used in diagnostic
 CC compositions and vaccines for mycobacteria other than of the
 CC M. tuberculosis complex, e.g. M. avium which infects poultry and they do
 CC not react with lymphoid cells previously primed with M. tuberculosis
 CC complex mycobacteria, and so do not give rise to a diagnostic reaction
 CC in individuals infected with these bacteria. The polypeptides may also
 CC be used in in vitro diagnostic tests e.g. stimulation of interferon-gamma
 CC release from lymphocytes. The polypeptide has similar or higher
 CC bioactivity as currently used tuberculin reagent in the standard
 CC delayed type hypersensitivity (DTB) skin test for tuberculosis, but may
 CC have greater specificity, being better able to discriminate between
 CC lymphoid cells primed from tuberculosis and from previous vaccination.
 CC The present sequence represents M. tuberculosis ESAT-6 used in the
 CC exemplification of the present invention.
 CC
 SQ Sequence 95 AA;
 QY
 Db 1 MTEOQWNPAGIEAANA 15
 |||||
 1 MTEOQWNPAGIEAANA 15
 RESULT 15
 AAY29788
 ID AAY29788 standard; Protein: 95 AA.
 XX
 AC AAY29788;
 XX
 DT 08-NOV-1999 (first entry)
 XX
 DE Mycobacterium tuberculosis ESAT-6 protein.
 XX
 KW Mycobacterial; ESAT-6; 6 kDa antigen; tuberculosis; infection;
 KW immunological response; diagnosis; vaccine.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN US5955077-A.
 XX
 PD 21-SEP-1999.

```

XX 05-JUN-1995; 95US-0465640.
PF
XX
PR 05-JUN-1995; 95US-0465640.
PR 20-SEP-1993; 93US-0123182.
PR 01-JUL-1994; 94WO-DK00273.
XX
PA (STAT-) STATENS SERUMINSTITUT.
XX
PI Andersen AB, Andersen P, Haslov K, Sorensen AL;
DR WPI; 1999-539545/45.
DR N-PSDB; AA208877.
XX
PT Polypeptide secreted from Mycobacterium is useful as a vaccine
PT against tuberculosis
XX
PS Claim 24; Fig 10C; 39pp; English.
XX
CC The present invention describes a purified or non-naturally occurring
CC polypeptide (I) released from a metabolising mycobacteria comprising an
CC ESAT6, also called the 6 kDa antigen. The present sequence represents
CC Mycobacterium tuberculosis ESAT-6 protein. Also described is a purified
CC or non-naturally occurring polypeptide (II) with the ability to elicit
CC a delayed type hypersensitivity reaction which comprises a T cell
CC epitope of (I). (II) can be used with a carrier or vehicle in a
CC composition for diagnosing tuberculosis caused by mycobacteria belonging
CC to the tuberculosis complex, i.e. Mycobacterium tuberculosis, M. bovis
CC and M. africanum. The composition can be used to detect microbial
CC antibodies or components of mycobacteria in samples or in animals
CC through the use of immunoassays. (II) can be used as a vaccine for
CC immunising an animal, including humans against tuberculosis caused by
CC mycobacteria of the tuberculosis-complex. (II) induce a release of
CC IFN-gamma from reactivated T-lymphocytes evoking a protective immune
CC response. Vaccine containing (II) has the same protective potency as
CC the live BCG vaccine against tuberculosis.
XX
SQ Sequence 95 AA;

```

Query Match 100.0%; Score 79; DB 20; Length 95;
 Best Local Similarity 100.0%; Pred. No. 8.1e-06;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MTEQQWNFAGIEAA 15
   |||||
Db 1 MTEQQWNFAGIEAA 15

```

Search completed: July 3, 2003, 14:05:31
 Job time : 31 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2003, 14:02:14 ; Search time 6.5 Seconds
(without alignments)
95.715 Million cell updates/sec

Title: US-09-830-839-1
Perfect score: 79
Sequence: 1 MTEGQWNFAGIEAA 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt.40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query | Match | Length | DB ID | Description |
|------------|-------|-------|-------|--------|--------------|---------------------|
| 1 | 74 | 93.7 | 94 | 1 | ESAB_MYCTU | Q57165 mycobacteri |
| 2 | 45 | 57.0 | 476 | 1 | NOEA_RHIME | Q52892 rhizobium m |
| 3 | 42 | 53.2 | 539 | 1 | G6PI_CAUCR | Q9abk5 caulobacter |
| 4 | 41 | 51.9 | 176 | 1 | DSBB_ECOLI | P30018 escherichia |
| 5 | 41 | 51.9 | 176 | 1 | DSBB_SHIFL | O54155 shigella fl |
| 6 | 38 | 48.1 | 305 | 1 | YNA5_DEIRA | Q91717 deinococcus |
| 7 | 38 | 48.1 | 438 | 1 | SGCE_HUMAN | O43556 homo sapien |
| 8 | 38 | 48.1 | 444 | 1 | CADB_ECOLI | P23891 escherichia |
| 9 | 38 | 48.1 | 508 | 1 | HMCN_PIG | O02734 sus scrofa |
| 10 | 38 | 48.1 | 552 | 1 | DREB_CHICK | P18302 gallus gall |
| 11 | 38 | 48.1 | 2564 | 1 | SPCO_HUMAN | Q98254 homo sapien |
| 12 | 37 | 46.8 | 211 | 1 | MT04_SYNT3 | P73161 synechocyst |
| 13 | 37 | 46.8 | 264 | 1 | COMT_RAT | P22734 rattus norv |
| 14 | 37 | 46.8 | 294 | 1 | G3P_ESCBL | P24749 escherichia |
| 15 | 37 | 46.8 | 335 | 1 | YD59_SCHHO | Q10312 schizosacch |
| 16 | 37 | 46.8 | 354 | 1 | WN11_BRARE | O73864 brachydanio |
| 17 | 37 | 46.8 | 445 | 1 | YJDE_ECOLI | P39269 escherichia |
| 18 | 37 | 46.8 | 475 | 1 | GUNA_CLOCE | P17901 clostridium |
| 19 | 37 | 46.8 | 668 | 1 | TM18_YEAST | Q04511 saccharomyc |
| 20 | 37 | 46.8 | 879 | 1 | STYL_XYLFA | Q9pb98 xyliella fas |
| 21 | 36.5 | 46.2 | 947 | 1 | LKTA_PASSP | P55123 pasteurella |
| 22 | 36 | 45.6 | 131 | 1 | IBN3_SHIDY | P03832 shigella dy |
| 23 | 36 | 45.6 | 148 | 1 | HPCR_ECOLI | O07095 escherichia |
| 24 | 36 | 45.6 | 175 | 1 | RP18_RAT | P47940 rattus norv |
| 25 | 36 | 45.6 | 230 | 1 | TERM_BPCP1 | Q37988 bacteriopho |
| 26 | 36 | 45.6 | 326 | 1 | ODPB_RICPR | Q9ad13 rickettsia |
| 27 | 36 | 45.6 | 430 | 1 | PYRC_LACPL | P77884 lactobacilli |
| 28 | 36 | 45.6 | 475 | 1 | ARCD_LACSP | O53092 lactobacilli |
| 29 | 36 | 45.6 | 496 | 1 | SRM_MOUSE | O62270 mus musculu |
| 30 | 36 | 45.6 | 499 | 1 | MURE_MOUSE | O84171 bruceella me |
| 31 | 36 | 45.6 | 511 | 1 | HUTH_RHIME | O31197 rhizobium m |
| 32 | 36 | 45.6 | 700 | 1 | FERG2_RALSTO | O84197 ralstonia s |
| 33 | 36 | 45.6 | 901 | 1 | SECA_HAEIN | P43803 haemophilus |

| | | | | | | |
|----|------|------|------|---|------------|--------------------|
| 34 | 36 | 45.6 | 964 | 1 | AMPN_RAT | P15684 rattus norv |
| 35 | 36 | 45.6 | 965 | 1 | AMPN_MOUSE | P97449 mus musculu |
| 36 | 36 | 45.6 | 994 | 1 | POLN_MIDDV | P03318 middsburg |
| 37 | 36 | 45.6 | 3106 | 1 | LMA2_MOUSE | O60675 mus musculu |
| 38 | 36 | 45.6 | 3110 | 1 | LMA2_HUMAN | P24043 homo sapien |
| 39 | 35.5 | 44.9 | 585 | 1 | PTNB_MOUSE | P35235 mus musculu |
| 40 | 35.5 | 44.9 | 593 | 1 | PTNB_HUMAN | O06124 homo sapien |
| 41 | 35.5 | 44.9 | 593 | 1 | PTNB_MOUSE | P41498 rattus norv |
| 42 | 35 | 44.3 | 74 | 1 | CKSL_DROME | Q24152 drosophila |
| 43 | 35 | 44.3 | 82 | 1 | Y4KO_RHISN | P55533 rhizobium s |
| 44 | 35 | 44.3 | 145 | 1 | Y008_MYCTU | P71577 mycobacteri |
| 45 | 35 | 44.3 | 175 | 1 | RP18_MOUSE | P47939 mus musculu |

ALIGNMENTS

RESULT 1
ID ESAB_MYCTU STANDARD: PRT: 94 AA.
AC Q57165; 084901;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 6 kDa early secretory antigenic target (ESAT-6).
GN ESAT6 OR RV3875 OR MT3989 OR MTV027.10.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
[1]
RN SEQUENCE FROM N.A.
RP SPECIES=M.tuberculosis; STRAIN=H37RV;
RC MEDLINE=95204931; PubMed=7897219;
RA Andersen P., Andersen A.B., Sorensen A.L., Nagai S.;
RT "Recall of long-lived immunity to Mycobacterium tuberculosis infection
in mice";
RL J. Immunol. 154:3359-3372(1995).
[2]
RN SEQUENCE FROM N.A., SEQUENCE OF 1-10, AND CHARACTERIZATION.
RP SPECIES=M.tuberculosis; STRAIN=Erdmann;
RC MEDLINE=95247251; PubMed=7729876;
RA Sorensen A.L., Nagai S., Houen G., Anderson P., Anderson A.B.;
RT "Purification and characterization of a low-molecular-mass T-cell
antigen secreted by Mycobacterium tuberculosis";
RL Infect. Immun. 63:1710-1717(1995).
[3]
RN SEQUENCE FROM N.A.
RP SPECIES=M.tuberculosis; STRAIN=H37RV;
RC MEDLINE=96295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
Bardock K., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
Davies R., Devlin K., Feltham D., Brown D., Chillingworth T., Connor R.,
Hornsey T., Jagels K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
Rutten S., Seeger K., Skelton S., Squares S., Squares R.,
Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence";
RL Nature 393:537-544(1998).
[4]
RN SEQUENCE FROM N.A.
RP SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RC Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
Peterson J.F., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
Kolonyak J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
Bisbal W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains";
RL submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

```

RN [5]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis;
RA Singh B., Siddiqui Z., Singh S., Sharma P.;
RT "ESAT-6 gene of a clinical isolate of Mycobacterium tuberculosis from
  India." (SEP-2001) to the EMBL/GenBank/DBJ databases.
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis;
RA Mahdara G.G., Sabo P.J., Hickey M.J., Singh D.C., Storer C.K.;
RT "Molecular analysis of genetic differences between Mycobacterium bovis
  BCG and virulent M. bovis."
RL J. Bacteriol. 178:1274-1282(1996).
RN [7]
RP SEQUENCE OF 1-70 FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37RV;
RA MEDLINE=96061212; PubMed=9846755;
RA Berthet F., X., Rasmussen P.B., Rosenkrands I., Andersen P.,
  Gicquel B.;
RT "A Mycobacterium tuberculosis operon encoding ESAT-6 and a novel
  low-molecular-mass culture filtrate protein (CFP-10)."
RL Microbiology 144:3195-3203(1998).
CC -1- FUNCTION: NOT KNOWN. ELICITS HIGH LEVEL OF INF-GAMMA FROM MEMORY
  EFFECTOR CELLS DURING THE FIRST PHASE OF A PROTECTIVE IMMUNE
  RESPONSE.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE ESAT6 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U34848; AAC44033.1; -
CC EMBL: X79562; CA56099.1; -
CC EMBL: AL022120; CA117967.1; -
CC EMBL: AF420491; AAL16896.1; -
CC EMBL: AE007190; AAK48357.1; -
CC EMBL: AF004671; AAC83446.1; -
CC TIGR: MT3989; -
CC Tuberculist: RV3875; -
CC Antigen: Complete proteome.
CC INIT_MET 0
CC FT 0
CC SEQUENCE 94 AA; 9773 MW; 19245B0EC478BC84 CRC64;
Query Match 93.7%; Score 74; DB 1; Length 94;
Best Local Similarity 100.0%; Pred. No. 5.7e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 TEQOMFAGIEAAA 15
DB 1 TEQOMFAGIEAAA 14

```

RESULT 2

```

NOEA_RHIME
ID NOEA_RHIME STANDARD; PRT; 476 AA.
AC 052892;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Modulation protein noea.
GN NOEA OR RA0416 OR SMA0773.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Plasmid pSYMA (megaplasmid 1).
OC Bacteria: Proteobacteria: alpha subdivision: Rhizobiaceae group;
OC Rhizobiaceae: Sinorhizobium.
OX NCBI_TaxID=382;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RCR2011 / SU47;
RX MEDLINE=96111489; PubMed=8801423;
RA Ardourel M., Lortet G., Maillet F., Roche P., Truchet G.,
  Promé J.-C., Rosenberg C.;
RT "In Rhizobium meliloti, the operon associated with the nod box n5
  comprises nodJ, noeA and noeB, three host-range genes specifically
  required for the nodulation of particular Medicago species."
RL Mol. Microbiol. 17:687-699(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
  Barloy-Hubler F., Bowser L., Capela D., Galbert F., Gouzy J.,
  Gurjel M., Hong A., Huzar L., Hyman R.W., Kahn D., Kahn M.L.,
  Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
  Yen K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
  Sinorhizobium meliloti pSYMA megaplasmid."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
CC -1- FUNCTION: NOT KNOWN. DOES NOT SEEM TO PARTICIPATE IN NOD FACTOR
  SYNTHESIS BUT REQUIRED FOR NODULATION ON SOME SPECIFIC MEDICAGO
  SPECIES SUCH AS M.LITTORALIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U26430; AAC44091.1; -
CC EMBL: AE007232; AAK65074.1; -
CC InterPro: IPR000051; SAM_bind.
CC Plasmid: Nodulation; Complete proteome.
CC SEQ 476 AA; 53673 MW; EF0A4009B5F4965E CRC64;
Query Match 57.0%; Score 45; DB 1; Length 476;
Best Local Similarity 63.6%; Pred. No. 2.6;
Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 5 QMNFAGIEAAA 15
DB 100 QMNFAGIEAAA 110

```

RESULT 3

```

G6PI_CAUCR
ID G6PI_CAUCR STANDARD; PRT; 539 AA.
AC 09ABK5;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glucose-6-phosphate isomerase (EC 5.3.1.9) (GPI) (Phosphoglucose
  isomerase) (PGI) (Phosphohexose isomerase) (PHI).
GN PGI OR CC0222.
OS Caulobacter crescentus.
OC Bacteria: Proteobacteria: alpha subdivision: Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
  Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
  Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
  Debay R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
  Kolonay J.F., Smit J., Craven M.B., Kouri H., Shetty J., Berry K.,
  Utterback T., Tian K., Wolf A., Vamathevan J., Ermolaeva M., White O.,

```


RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RT "Complete genome sequence of *Caulobacter crescentus*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 CC -1- CATALYTIC ACTIVITY: D-glucose 6-phosphate - D-fructose 6-phosphate.
 CC -1- PATHWAY: Involved in glycolysis and in gluconeogenesis.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE GPI FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL: AE005696; AAK22209.1; -
 DR HSSP: Q9N1E2; IHOX.
 DR TIGR: CC0222; -
 DR InterPro: IPR001672; G6P_Isomerase.
 DR Pfam: PF00342; PGI.1.
 DR PRINTS: PR00662; G6PISOMERASE.
 DR PROSITE: PS00765; P-GLUCOSE-ISOMERASE.1; 1.
 DR PROSITE: PS00174; P-GLUCOSE-ISOMERASE.2; 1.
 KM Isomerase: Gluconeogenesis; Glycolysis; Complete proteome.
 FT ACT_SITE 380 380 By SIMILARITY.
 FT ACT_SITE 508 508 By SIMILARITY.
 SQ SEQUENCE 539 AA; 57526 MW; B233DFBAF7FD1595 CRC64;

 Query Match 53.2%; Score 42; DB 1; Length 539;
 Best Local Similarity 50.0%; Pred. No. 9.6;
 Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

 QY 1 MTEQNNFACIEAA 14
 Db 48 LSKQAWDEAGLEAA 61

 RESULT 4
 DSB_ECOLI STANDARD; PRT; 176 AA.
 AC P30018;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Disulfide bond formation protein B (Disulfide oxidoreductase).
 GN DSB OR ROXB OR B185 OR Z1948 OR ECS1680.
 OS *Escherichia coli*, and
 OS *Escherichia coli* O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC *Escherichia*.
 OX NCBI_Taxid=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
 RX MEDLINE=93157338; PubMed=8430071.
 RA Bardwell J.C.A., Lee J.-O., Jander G., Martin N., Belin D.,
 RA Beckwith J.;
 RT "A pathway for disulfide bond formation in vivo."
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1038-1042(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93348217; PubMed=7688471;
 RA Masiakos D., Georgopoulos C., Raina S.;
 RT "Identification and characterization of the *Escherichia coli* gene
 RT *dsbB*, whose product is involved in the formation of disulfide bonds
 RT in vivo."
 RL Proc. Natl. Acad. Sci. U.S.A. 90:7084-7088(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX STRAIN-K12 / MG1655;
 RX MEDLINE=9746617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12."
 RL science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiiuchi T.;
 RT "A 7.18-kb DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map."
 RL DNA Res. 3:137-155(1996).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lam A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7."
 RL Nature 409:529-533(2001).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasaki A., Ogasawara N., Yasunaga T.,
 RA Kubara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
 RT O157:H7 and genomic comparison with a laboratory strain K-12."
 RL DNA Res. 8:11-22(2001).
 RN [7]
 RP PRELIMINARY SEQUENCE OF 1-169 FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=92283803; PubMed=1317851;
 RA Plummer E., Padan E., Schudiner S.;
 RT "Cloning, sequencing, and expression of the *nhaB* gene, encoding a
 RT Na⁺/H⁺ antiporter in *Escherichia coli*."
 RL J. Biol. Chem. 267:11064-11068(1992).
 RN [8]
 RP TOPOLOGY AND MUTAGENESIS OF CYSTEINE RESIDUES.
 RX MEDLINE=95045404; PubMed=7957076;
 RA Jander G., Martin N.L., Beckwith J.;
 RT "Two cysteines in each periplasmic domain of the membrane protein
 RT DsbB are required for its function in protein disulfide bond
 RT formation."
 RL EMBO J. 13:5121-5127(1994).
 RN [9]
 RP ACTIVE SITES.
 RX MEDLINE=99164086; PubMed=11064586;
 RA Kobayashi T., Ito K.;
 RT "Respiratory chain strongly oxidizes the CXXC motif of DsbB in the
 RT *Escherichia coli* disulfide bond formation pathway."
 RL EMBO J. 18:1192-1198(1999).
 CC -1- FUNCTION: REQUIRED FOR DISULFIDE BOND FORMATION IN SOME
 CC PERIPLASMIC PROTEINS SUCH AS PHO A OR OMP A. ACTS BY OXIDIZING
 CC THE DSB A PROTEIN.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
 CC -1- SIMILARITY: BELONGS TO THE DSB FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: L03721; AAA3711.1; ALT_INIT.
 DR EMBL: AE000216; AAC74269.1; -
 DR EMBL: D90752; BAA36032.1; ALT_INIT.
 DR EMBL: D90753; BAA36040.1; ALT_INIT.
 DR EMBL: AE005335; AAG56036.1; -
 DR EMBL: AP002555; BAB35103.1; -
 DR EMBL: M83655; -; NOT_ANNOTATED_CDS.
 DR PIR: JCI109; JCI109.
 DR Ecocore: EG11393; dsdb.
 DR InterPro: IPR003752; dsdb.
 DR Pfam: PF02600; dsdb; 1.
 DR Oxidoreductase; Redox-active center; Electron transport; Chaparone;
 KM Transmembrane; Inner membrane; Complete proteome.
 FT DOMAIN 1 14
 FT TRANSMEM 15 31
 FT DOMAIN 32 49
 FT TRANSMEM 50 65
 FT DOMAIN 66 71
 FT TRANSMEM 72 89
 FT DOMAIN 90 144
 FT TRANSMEM 145 163
 FT DOMAIN 164 176
 FT DISULFID 41 44
 FT DISULFID 104 130
 FT CONFLICT 50 51
 FT CONFLICT 136 136
 SQ SEQUENCE 176 AA; 2014 MW; 9C8D673D51E9F09B CRC64;

Query Match 51.9%; Score 41; DB 1; Length 176;
 Best Local Similarity 60.0%; Pred. No. 4.7;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 3 ECOMNFAGIE 12
 1:11:11:11

DB 132 ERQWDFLGIE 141

RESULT 5
 DSB: SHIFL STANDARD; PRT; 176 AA.
 ID Q54155;
 AC 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Disulfide bond formation protein B (Disulfide oxidoreductase).
 GN DSB.
 OS *Shigella flexneri*.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC *Shigella*.
 OX NCBI_TaxID=623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ISH600 / Serotype 2a;
 RA Sasakawa C.;
 RL Submitted (Aug-1995) to the EMBL/Genbank/DBA databases.
 CC -1- FUNCTION: REQUIRED FOR DISULFIDE BOND FORMATION IN SOME
 CC PERIPLASMIC PROTEINS SUCH AS PHO A OR OMPA. ACTS BY OXIDIZING
 CC THE DSB PROTEIN (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE DSB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC or send an email to license@isb-sib.ch).

DR EMBL: D38254; BAA07408.1; -
 DR InterPro: IPR003752; Dsdb.
 DR Pfam: PF02600; Dsdb; 1.
 DR Oxidoreductase; Redox-active center; Electron transport; Chaparone;
 KM Transmembrane; Inner membrane.
 FT DOMAIN 1 14
 FT TRANSMEM 15 31
 FT DOMAIN 32 49
 FT TRANSMEM 50 65
 FT DOMAIN 66 71
 FT TRANSMEM 72 89
 FT DOMAIN 90 144
 FT TRANSMEM 145 162
 FT TRANSMEM 163 176
 FT DISULFID 41 44
 FT DISULFID 104 130
 SQ SEQUENCE 176 AA; 2003 MW; 7552F2B916B72246 CRC64;

Query Match 51.9%; Score 41; DB 1; Length 176;
 Best Local Similarity 60.0%; Pred. No. 4.7;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 3 ECOMNFAGIE 12
 1:11:11:11

DB 132 ERQWDFLGIE 141

RESULT 6
 YN45_DEIRA STANDARD; PRT; 305 AA.
 ID YN45_DEIRA
 AC Q9RRY7;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein DR2345.
 GN DR2345.
 OS *Deinococcus radiodurans*.
 OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
 OC Deinococcaceae; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RI;
 RX MEDLINE=20036896; PubMed=10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Uyang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.T., Lam P., McDonald L., Uitterback T., Zalewski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RA "Genome sequence of the radioresistant bacterium Deinococcus
 RT radiodurans RI".
 RL Science 286:1571-1577(1999).
 CC -1- SIMILARITY: BELONGS TO THE UPF0151 FAMILY. SOME SIMILARITY TO
 CC PHOSPHATASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AE002065; AAF11890.1; -
 DR TIGR: DR2345; -
 DR InterPro: IPR004843; M-peptidase.
 DR InterPro: IPR004844; S/T phosphatase.
 DR Pfam: PF00149; Metallophos; 1
 KW Hypothetical protein; Hydrolase; Complete proteome.

SQL SEQUENCE 305 AA; 32238 MW; FBF59ADE18CA0F88 CRC64;
Query Match 48.1%; Score 38; DB 1; Length 305;
Best Local Similarity 45.5%; Pred. No. 26;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
OY 1 MTEQOWNFAGI 11
1 : 111 :
47 VTHEQOWNLVGL 57
DB
RESULT 7
SCCE_HUMAN STANDARD: PRT; 438 AA.
AC 043556;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epsilon-sarcoglycan precursor (Epsilon-Sg).
GN SCCE OR ESG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RA Nigro V.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RA McNally E.M., Ly C.T., Kunkel L.M.;
RX MEDLINE=98133310; PubMed=9475163;
RT "Human epsilon-sarcoglycan is highly related to alpha-sarcoglycan (adhalin), the limb girdle muscular dystrophy 2D gene.";
RL FEBS Lett. 422:27-32(1998).
RN [3]
RP SEQUENCE OF 269-413 FROM N.A.
RX MEDLINE=98070432; PubMed=9405466;
RA Ettlinger A.J., Feng G., Sanes J.R.;
RT "Epsilon-sarcoglycan, a broadly expressed homologue of the gene mutated in limb girdle muscular dystrophy 2D.";
RL J. Biol. Chem. 272:32534-32538(1997).
RN [4]
RP ERRATUM.
RA Ettlinger A.J., Feng G., Sanes J.R.;
RL J. Biol. Chem. 273:19922-19922(1998).
CC -1- FUNCTION: COMPONENT OF THE SARCOGLYCAN COMPLEX, A SUBCOMPLEX OF THE DYSTROPHIN-GLYCOPROTEIN COMPLEX WHICH FORMS A LINK BETWEEN THE F-ACTIN CYTOSKELETON AND THE EXTRACELLULAR MATRIX.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Sarcolemmal (Potential).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; SCCE-1 (SHOWN HERE) AND SCCE-2; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: UBQUITOUS.
CC -1- SIMILARITY: BELONGS TO THE SARCOGLYCAN ALPHA/EPSILON FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to license@sdb-sdb.ch).
CC -----
CC DR EMBL; AF031920; AAC14021.1; ALT_INIT.
CC DR EMBL; AF036364; AAC04368.1; -
CC DR EMBL; AJ000534; CAA04167.1; -
CC DR Genbank; HGNC:10808; SCCE.
CC MIM; 604149; -
CC Cytoskeleton; Transmembrane; Glycoprotein; Signal;
KW Alternative splicing.

FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 438 EPSILON-SARCOGLYCAN.
FT DOMAIN 23 293 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 294 314 POTENTIAL.
FT DOMAIN 315 438 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 211 317 CYS-RICH.
FT CARBOHYD 176 176 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 395 419 MISSING (IN ISOFORM SCCE-2).
SQ SEQUENCE 438 AA; 49754 MW; D87D189959C16F23 CRC64;
OY 1 MTEQOWSFAPVAQA 14
1 : 111 : 1
390 MTEQOWSFAPVAQA 403
DB
Query Match 48.1%; Score 38; DB 1; Length 438;
Best Local Similarity 50.0%; Pred. No. 38;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
RESULT 8
CADA_ECOLI STANDARD: PRT; 444 AA.
ID CADA_ECOLI
AC P23691;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable cadaverine/lysine antiporter.
GN CADA OR BA132 OR Z5735 OR ECSS114.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92105022; PubMed=1370290;
RA Watson N., Dunyak D.S., Rosey E.L., Slonczewski J.L., Olson E.R.;
RT "Identification of elements involved in transcriptional regulation of the Escherichia coli cad operon by external pH.";
RL J. Bacteriol. 174:530-540(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92210511; PubMed=1556085;
RA Meng S.-Y., Bennett G.N.;
RT "Nucleotide sequence of the Escherichia coli cad operon: a system for neutralization of low extracellular pH.";
RL J. Bacteriol. 174:2659-2669(1992).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RA MEDLINE=9534362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.T., Daniels D.L., Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=O157:H7 / EDJ933 / ATCC 700927;
RA MEDLINE=21074935; PubMed=11206551;
RA Perra N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D., Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A., Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L., Grodzick E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocostas K., Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C., Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [5]
RP SEQUENCE FROM N.A.
RX STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,

RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kohara S., Shiba T., Hattori M., Shlnagawa H.,
 RT Complete genome sequence of enterohemorrhagic *Escherichia coli*
 RT O157:H7 and genomic comparison with a laboratory strain K-12.*;
 RL DNA Res. 8:11-22(2001).
 CC -1- FUNCTION: PROBABLE CADYERINE/LYSINE ANTI-PORTER OR PART OF IT.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
 CC -1- INDUCTION: AT LOW PH IN THE PRESENCE OF LYSINE AND ACETEVES
 CC MAXIMAL LEVEL UNDER ANAEROBIC CONDITIONS.
 CC -1- SIMILARITY: BELONGS TO THE ARCD/CADB/GERAB/LYSI FAMILY OF
 CC PERMEASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M67452; AAA23532.1; -;
 DR EMBL: M76411; AAA23535.1; -;
 DR EMBL: U14003; AAA97032.1; -;
 DR EMBL: AE000486; AAC77093.1; -;
 DR EMBL: AE005647; AAG59332.1; -;
 DR EMBL: AF002568; BAB38537.1; -;
 DR PIR: B41968; B41968.
 DR PIR: A41842; A41842.
 DR Ecogene: EG10132; CADB.
 DR InterPro: IPR002293; AA/tel_primease1.
 DR InterPro: IPR004841; Permease.
 DR Pfam: PF00324; aa_permeases; 1.
 DR TrGFams: TRGR00905; 2A0302; 1.
 KW Transport; Antiport; Amino-acid transport; Transmembrane;
 KW Inner membrane; Complete proteome.
 FT TRANSMEM 7 27 POTENTIAL.
 FT TRANSMEM 35 55 POTENTIAL.
 FT TRANSMEM 68 88 POTENTIAL.
 FT TRANSMEM 95 115 POTENTIAL.
 FT TRANSMEM 123 143 POTENTIAL.
 FT TRANSMEM 149 169 POTENTIAL.
 FT TRANSMEM 193 213 POTENTIAL.
 FT TRANSMEM 222 242 POTENTIAL.
 FT TRANSMEM 250 270 POTENTIAL.
 FT TRANSMEM 273 293 POTENTIAL.
 FT TRANSMEM 323 343 POTENTIAL.
 FT TRANSMEM 354 374 POTENTIAL.
 FT TRANSMEM 384 404 POTENTIAL.
 FT TRANSMEM 405 425 POTENTIAL.
 SQ SEQUENCE 444 AA; 46665 MW; E87913B449B0500A CRC64;
 Query Match 48.1%; Score 38; DB 1; Length 444;
 Best Local Similarity 60.0%; Pred. No. 38;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 6 WNFAGIEAIA 15
 DB 198 MAEFGVESAA 207
 RESULT 9
 HMCM_PIG STANDARD; PRT; 508 AA.
 AC 002734;
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Hydroxymethylglutaryl-CoA synthase, mitochondrial precursor
 DE (EC 4.1.3.5) (HMG-CoA synthase) (3-hydroxy-3-methylglutaryl coenzyme A
 DE synthase).
 GN HMGCS2.

OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Liver;
 RX MEDLINE=97293201; PubMed=9164842;
 RA Adams S.H., Albo C.S., Asins G., Hegardt F.G., Marrero P.F.;
 RT 'Gene expression of mitochondrial 3-hydroxy-3-methylglutaryl-CoA
 RT synthase in a poorly ketogenic, mammal: effect of starvation during
 RT the neonatal period of the piglet.';
 RL Biochem. J. 324:65-73(1997).
 CC -1- FUNCTION: THIS ENZYME CONDENSES ACETYL-CoA WITH ACETOACETYL-CoA
 CC TO FORM HMG-CoA, WHICH IS THE SUBSTRATE FOR HMG-CoA REDUCTASE.
 CC -1- CATALYTIC ACTIVITY: (S)-3-hydroxy-3-methylglutaryl-CoA + CoA -
 CC acetyl-CoA + H(2)O + acetoacetyl-CoA.
 CC -1- PATHWAY: RESPONSIBLE, TOGETHER WITH HMG-CoA LYASE FOR KETONE
 CC BODY BIOSYNTHESIS.
 CC -1- SUBCELLULAR LOCATION: Mitochondrial.
 CC -1- SIMILARITY: BELONGS TO THE HMG-CoA SYNTHASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U90884; AAC48727.1; -;
 DR InterPro: IPR000590; HMG-CoA_synth.
 DR Pfam: PF01154; HMG-CoA_synth; 1.
 DR PROSITE: PS01226; HMG-CoA SYNTHASE; 1.
 KW lyase; Cholesterol biosynthesis; Mitochondrion; Transit peptide;
 KW Multigene family.
 FT TRANSMEM 1 37 POTENTIAL.
 FT CHAIN 38 508 HYDROXYMETHYLGUTARYL-CoA SYNTHASE.
 FT ACT_SITE 166 166 POTENTIAL.
 SQ SEQUENCE 508 AA; 56933 MW; 5479DE6F70B3C0F6 CRC64;
 Query Match 48.1%; Score 38; DB 1; Length 508;
 Best Local Similarity 70.0%; Pred. No. 44;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 3 EDOMFAGIE 12
 DB 278 EKWMKQAGIE 287
 RESULT 10
 DREB_CHICK STANDARD; PRT; 652 AA.
 AC P18302; Q91358; Q91359;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Drebrin (Developmentally regulated brain protein).
 GN DBN1.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS E1 AND E2).
 RX TISSUE=Brain;
 RX MEDLINE=89089137; PubMed=3208110;
 RA Kojima N., Kato Y., Shitao T., Obata K.;
 RT Nucleotide sequences of two embryonic drebrins, developmentally
 RT regulated brain proteins, and developmental change in their mRNAs.;
 RL Brain Res. 464:207-215(1988).
 RN [2]

RP SEQUENCE FROM N.A. (ISOFORM A).
RA MEDLINE-93368392; PubMed-8361332;
RX Kojima N., Shirata T., Obata K.;
RT "Molecular cloning of a developmentally regulated brain protein,
RT chicken drebrin A and its expression by alternative splicing of the
RT drebrin gene.";
RL Brain Res. Mol. Brain Res. 19:101-114(1993).
CC -1- FUNCTION: DREBRIN MIGHT PLAY SOME ROLE IN CELL MIGRATION,
CC EXTENSION OF NEURONAL PROCESSES AND PLASTICITY OF DENDRITES,
CC RESPECTIVELY. BINDS F-ACTIN.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- ALTERNATIVE PRODUCTS: 3 isoforms: A (shown here), E1 and E2;
CC are produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: BRAIN NEURONS.
CC -1- MISCELLANEOUS: DREBRIN ARE CLASSIFIED INTO TWO FORMS OF THE
CC EMBRYONIC TYPE (E1 AND E2) AND ONE FORM OF THE ADULT TYPE (A). THE
CC TIME COURSE OF THEIR APPEARANCE ARE DIFFERENT FROM EACH OTHER.
CC THEIR STRUCTURES ARE CLOSELY RELATED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M36961; AAA8750.1; -
DR EMBL: S65296; AAB28012.1; JOINED.
DR EMBL: S65297; AAB28012.1; JOINED.
DR EMBL: S65280; AAB28012.1; JOINED.
DR EMBL: S65281; AAB28012.1; JOINED.
DR EMBL: S65288; AAB28012.1; JOINED.
DR EMBL: S65289; AAB28012.1; JOINED.
DR EMBL: S65290; AAB28012.1; JOINED.
DR EMBL: S65291; AAB28012.1; JOINED.
DR EMBL: S65292; AAB28012.1; JOINED.
DR EMBL: S65294; AAB28012.1; JOINED.
DR EMBL: S65267; AAB28011.2; -
DR EMBL: S65230; AAB28010.1; -
DR PIR: A43776; A43776.
DR InterPro: IPR002108; Actbind_cofln.
DR Pfam: PF00241; coflin_ADF.1.
DR SMART: SM00102; ADF.1.
KW Actin-binding; Neurone; Alternative splicing.
FT DOMAIN 47 50 POLY-GLY.
FT DOMAIN 332 339 POLY-SER.
FT DOMAIN 542 552 POLY-GLU.
FT VARSPPLIC 317 362 MISSING (IN ISOFORM E1).
FT VARSPPLIC 307 405 MISSING (IN ISOFORM E2).
FT CONFLICT 552 552 E -> EE (IN REF. 1).
SQ SEQUENCE 652 AA: 71535 MW: 8672CA549833E65 CRC64.
Query Match 48.1%; Score 38; DB 1; Length 652;
Best Local Similarity 57.1%; Pred. No. 56;
Matches 8; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Oy 2 TEOQMNPAIGIFAAA 15
Db 419 TEOHMPFGPEDKA 432
RESULT 11
SPCO_HUMAN STANDARD. PRT; 2564 AA.
AC Q9H254; Q9H3G8; Q9H1K7; Q9H1K8; Q9H1K9;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Spectrin beta chain, brain 3 (Spectrin, non-erythroid beta chain 3)
GN (beta-IV spectrin).
GN SPTBN4 OR SPTBN3 OR KIAA1642.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBT_Taxid-9606;
RN
RN SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE-21316449; PubMed-11294830;
RA Tse W.T., Tang J., Jin O., Korsgren C., John K.M., Kung A.L.,
RA Gwynn B., Peters L.L., Lux S.E.;
RT "A new spectrin, beta-IV, has a major truncated isoform that
RT associates with promyelocytic leukemia protein nuclear bodies and the
RT nuclear matrix.";
RL J. Biol. Chem. 276:23974-23985(2001).
[2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 TO 4).
RX MEDLINE-20539976; PubMed-11086001.
RA Begins S., Aguijaro D., Dirks R. Jr., Maksimova E., Stabach P.,
RA Hermet J.-M., Zhang J.-P., Philbrick W., Stepnev V., Ort T.,
RA Solimena M.;
RT "BetaIV spectrin, a new spectrin localized at axon initial segments
RT and nodes of ranvier in the central and peripheral nervous system.";
RL J. Cell Biol. 151:985-1002(2000).
[3]
RN SEQUENCE OF 386-2382 FROM N.A. (ISOFORM 1).
RP TISSUE=Brain;
RC MEDLINE-20450683; PubMed-1097877;
RX Nagase T., Kikuno R., Nakayama M., Hirosewa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.
RT XVII. The complete sequences of 100 new cDNA clones from brain which
RT code for large proteins in vitro.";
RL DNA Res. 7:273-281(2000).
CC -1- ALTERNATIVE PRODUCTS: 4 isoforms: 1 (shown here), 2, 3 and 4; are
CC produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: Abundantly expressed in brain and pancreatic
CC islets.
CC -1- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.
CC -1- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: CONTAINS 18 SPECTRIN REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF311855; AAG42473.1; -
DR EMBL: AF082075; AAG38874.1; -
DR EMBL: AY004226; AAF93171.1; -
DR EMBL: AY004226; AAF93172.1; -
DR EMBL: AY004227; AAF93173.1; -
DR EMBL: AB046862; BAB13468.1; -
DR HSSP: Q01082; 1BKR.
DR Genew: HGNC:14896; SPTBN4.
DR MIM: 606214; -
DR InterPro: IPR001589; Actbind_actnin.
DR InterPro: IPR001715; Calponin-like.
DR InterPro: IPR001849; PH.
DR InterPro: IPR002017; Spectrin.
DR InterPro: IPR001605; Spectrin_PH.
DR Pfam: PF00169; PH.3.
DR Pfam: PF00307; CH.2.
DR Pfam: PF00435; Spectrin.56.
DR PRINTS: PR00683; SPECTRINPH.
DR SMART: SM00033; CH.2.
DR SMART: SM00233; PH.1.
DR SMART: SM00150; SPEC.16.
DR PROSITE: PS00019; ACTININ_1.
DR PROSITE: PS00020; ACTININ_2.
DR PROSITE: PS00021; CH.2.
DR PROSITE: PS50003; PH_DOMAIN.1.

RA Vldgren J., Svensson L.A., Liljas A.;
 RT "Crystal structure of catechol O-methyltransferase.";
 CC Nature 368:354-358(1994).
 CC -1- FUNCTION: CATALYZES THE O-METHYLATION, AND THEREBY THE
 CC INACTIVATION, OF CATECHOLAMINE NEUROTRANSMITTERS AND CATECHOL
 CC HORMONES. ALSO SHORTENS THE BIOLOGICAL HALF-LIVES OF CERTAIN
 CC NEUROACTIVE DRUGS, LIKE L-DOPA, ALPHA-METHYL DOPA AND
 CC ISOPROTERENOL.
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + catechol = S-
 CC adenosyl-L-homocysteine + guaiacol.
 CC -1- COFACTOR: MAGNESIUM.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (isoform S-COMT). Type II
 CC membrane protein (isoform MB-COMT).
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; a membrane-bound form/MB-COMT
 CC (shown here) and a soluble form/S-COMT; are produced by
 CC alternative initiation.
 CC -1- PTM: The N-terminus is blocked.
 CC -1- SIMILARITY: TO OTHER MAMMALIAN CATECHOL-O-METHYLTRANSFERASE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z12651; GAA78276.1; -
 DR EMBL: M60754; AAA40882.1; ALT_INIT.
 DR EMBL: M60753; AAA40881.1; ALT_INIT.
 DR PIR: J00787; J00787.
 DR PIR: S22090; S22090.
 DR PDB: 1JYD; 1JYD-96.
 DR InterPro: IPR002935; Methyltransferase_3.
 DR InterPro: IPR000051; SAM_bind.
 DR Pfam: PF01596; Methyltransferase_3; 1.
 DR Transferrase; Methyltransferase; Neurotransmitter degradation;
 KW Catecholamine metabolism; Transmembrane; Signal-anchor; Magnesium;
 KW Alternative initiation; 3D-structure.
 FT CHAIN 1 264 CATECHOL-O-METHYLTRANSFERASE, MEMBRANE-
 FT BOUND ISOFORM.
 FT CHAIN 45 264 CATECHOL-O-METHYLTRANSFERASE, SOLUBLE
 FT INIT_MET 44 44 FOR SOLUBLE ISOFORM.
 FT TRANSMEM 3 19 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT ACT_SITE 187 187
 FT ACT_SITE 242 242
 FT METAL 184 184 MAGNESIUM.
 FT METAL 212 212 MAGNESIUM.
 FT METAL 213 213 MAGNESIUM.
 SQ SEQUENCE 264 AA; 29597 MW; F535DFE49C062854 CRC64;
 Query Match 46.8%; Score 37; DB 1; Length 264;
 Best Local Similarity 50.0%; Pred. No. 34;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 Oy 1 MTEOONFAGTE 12
 Db 141 ITQOMLNFRGLQ 152
 RESULT 14
 G3P_ESCBL STANDARD: PRT: 294 AA.
 AC P24749;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.12) (GAPDH)
 DE (Fragment).
 GN GAP.
 OS Escherichia blattae.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=563;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 29907, ATCC 33429, and ATCC 33430;
 RX MEDLINE=92065252; PubMed=1955870;
 RA Lawrence J.G., Ochman H., Harl D.L.;
 RT "Molecular and evolutionary relationships among enteric bacteria.";
 CC J. Gen. Microbiol. 137:1911-1921(1991).
 CC -1- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate + phosphate +
 CC NAD(+) = 3-phospho-D-glyceroyl phosphate + NADH.
 CC -1- PATHWAY: Second phase of glycolysis; first step.
 CC -1- SUBUNIT: HOMOTETRAMER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE
 CC DEHYDROGENASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M63358; AAA23837.1; -
 DR EMBL: M63359; AAA23852.1; -
 DR EMBL: M63360; AAA23855.1; -
 DR HSSP: P06977; IGAD.
 DR InterPro: IPR00173; GAP_dhdrogenase.
 DR Pfam: PF00044; gpdh; 1.
 DR Pfam: PF02800; gpdh.C; 1.
 DR PROSITE: PS00717; GAPDH; 1.
 KW Glycolysis; Oxidoreductase; NAD.
 FT NON_TER 1 1
 FT BINDING 135 135 GLYCERALDEHYDE 3-PHOSPHATE.
 FT ACT_SITE 162 162 ACTIVATES THIOL GROUP DURING CATALYSIS.
 FT NON_TER 294 294
 SQ SEQUENCE 294 AA; 31372 MW; 18562CB398177D8B CRC64;
 Query Match 46.8%; Score 37; DB 1; Length 294;
 Best Local Similarity 54.5%; Pred. No. 38;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 Oy 5 OMNFRGIRAA 15
 Db 69 KWNAGVEVVA 79
 RESULT 15
 YD59_SCHPO STANDARD: PRT: 335 AA.
 ID YD59_SCHPO
 AC Q10312; Q9YUL8;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein C6C3.09 in chromosome I.
 GN SPAC6C3.09.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser G.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson A.,

RA Holroyd S., Hornsby T., Howarth S., Hucke E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Welfjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Medler H., Wandutt R., Fumelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerretti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Siparovsky G.V., Ussery D., Barrell B.G., Nurse P.,
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z69731; CABA0281.1; -
KW Hypothetical protein.
SQ SEQUENCE 335 AA; 37948 MW; C1CD4E5850BAD323 CRC64;

Query Match 46.8%; Score 37; DB 1; Length 335;
Best Local Similarity 62.5%; Pred. NO. 43;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 6 WNFAGIEA 13
||| |:
Db 290 WNFAGIES 297

Search completed: July 3, 2003, 14:05:51
Job time : 7.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2003, 14:03:50 ; Search time 12.5 Seconds
(without alignments)
115.361 Million cell updates/sec

Title: US-09-830-839-1

Perfect score: 79

Sequence: 1 MTEQOMNFAGIEAANA 15

Scoring table: BIOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being predicted,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 79 | 100.0 | 95 | 2 A70803 | early secretory an |
| 2 | 45 | 57.0 | 476 | 2 S71360 | noea protein - rhi |
| 3 | 45 | 57.0 | 476 | 2 H85313 | NOEA host specific |
| 4 | 43 | 54.4 | 217 | 2 A82491 | fine protein VCA01 |
| 5 | 43 | 54.4 | 247 | 2 H90825 | hypothetical prote |
| 6 | 42 | 53.2 | 369 | 2 E86554 | Fe-S oxidoreductas |
| 7 | 42 | 53.2 | 672 | 2 G72069 | conserved hypothet |
| 8 | 42 | 53.2 | 539 | 2 E87276 | glucose-6-phosphat |
| 9 | 42 | 53.2 | 723 | 2 E87706 | prolyl oligopeptid |
| 10 | 41 | 51.9 | 176 | 1 F64864 | protein-disulfide |
| 11 | 41 | 51.9 | 176 | 2 H90838 | protein-disulfide |
| 12 | 41 | 51.9 | 176 | 2 H85696 | hypothetical prote |
| 13 | 41 | 51.9 | 445 | 2 D82342 | probable cadaverin |
| 14 | 41 | 51.9 | 471 | 2 E69435 | hypothetical prote |
| 15 | 40 | 50.6 | 38 | 2 H82256 | hypothetical prote |
| 16 | 39 | 49.4 | 219 | 2 AG2036 | hypothetical prote |
| 17 | 39 | 49.4 | 328 | 2 A70145 | prolipoprotein dia |
| 18 | 39 | 49.4 | 393 | 2 AH2150 | hypothetical prote |
| 19 | 39 | 49.4 | 714 | 2 T01762 | hypothetical prote |
| 20 | 39 | 49.4 | 714 | 2 G86844 | hypothetical prote |
| 21 | 39 | 49.4 | 916 | 2 D83093 | secretion protein |
| 22 | 38.5 | 48.7 | 392 | 2 A87572 | aminotransferase, |
| 23 | 38 | 48.1 | 252 | 2 PC1140 | cellulase (EC 3.2. |
| 24 | 38 | 48.1 | 284 | 2 E82868 | hypothetical prote |
| 25 | 38 | 48.1 | 297 | 2 AD2851 | cytochrome c1 floc |
| 26 | 38 | 48.1 | 305 | 1 H75285 | probable phosphos |
| 27 | 38 | 48.1 | 319 | 2 B97628 | cytochrome c1 (AFI |
| 28 | 38 | 48.1 | 443 | 2 AE0826 | probable cadaverin |
| 29 | 38 | 48.1 | 444 | 2 A41842 | lysine/cadaverine |

| | | | | | |
|----|----|------|-----|----------|----------------------|
| 30 | 38 | 48.1 | 444 | 2 B91268 | transport protein |
| 31 | 38 | 48.1 | 444 | 2 H86108 | transport of lysin |
| 32 | 38 | 48.1 | 593 | 2 I51213 | trehalin - chick |
| 33 | 38 | 48.1 | 607 | 2 A43776 | trehalin B2 - chick |
| 34 | 37 | 46.8 | 61 | 2 E86084 | hypothetical prote |
| 35 | 37 | 46.8 | 192 | 1 S75273 | hypothetical prote |
| 36 | 37 | 46.8 | 229 | 2 D90002 | hypothetical prote |
| 37 | 37 | 46.8 | 251 | 2 C81816 | hypothetical prote |
| 38 | 37 | 46.8 | 256 | 2 AD1115 | hypothetical prote |
| 39 | 37 | 46.8 | 257 | 2 AE1476 | hypothetical prote |
| 40 | 37 | 46.8 | 264 | 2 S22090 | catechol O-methylt |
| 41 | 37 | 46.8 | 267 | 2 A87404 | ribosomal protein |
| 42 | 37 | 46.8 | 294 | 2 I41220 | glycerolaldehyde-3-p |
| 43 | 37 | 46.8 | 326 | 2 A41862 | C-5 lyase (OHR326) |
| 44 | 37 | 46.8 | 335 | 2 T39033 | hypothetical prote |
| 45 | 37 | 46.8 | 348 | 2 A12475 | hypothetical prote |

ALIGNMENTS

```
RESULT 1
A70803
early secretory antigen target esat6 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: A70803 $49174
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajadream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MIMD:96295587; PMID:9634230
A:Accession: A70803
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-95 <COL>
A:Cross-references: GB:A1022120; GB:AL123456; NID:g3261558; PIDN:CA117967.1; PID:el26
A:Experimental source: strain H37RV
R:Soerensen, A.L.; Nagai, S.; Houen, G.; Andersen, P.; Andersen, A.B.
submitted to the EMBL Data Library, June 1994
A:Reference number: S49174
A:Accession: S49174
A:Molecule type: DNA
A:Residues: 1-13,'R',15-22,'S',24-95 <SOE>
A:Cross-references: EMBL:X79562; NID:g531708
C:Genetics:
A:Gene: esat6

Query Match          100.0%  Score 79;  DB 2;  Length 95;
Best Local Similarity 100.0%;  Pred. NO. 1.3e-06;
Matches 15;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY      1 MTEQOMNFAGIEAANA 15
DB      1 MTEQOMNFAGIEAANA 15

RESULT 2
S71360
noea protein - Rhizobium meliloti
C:Species: Rhizobium meliloti
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 04-Mar-2000
C:Accession: S71360; S71358
R:Ardoon, M.; Lortet, G.; Maillet, F.; Roche, P.; Truchet, G.; Prome, J.C.; Rosebn
submitted to the EMBL Data Library, May 1995
A:Reference number: S71360
A:Accession: S71360
A:Molecule type: DNA
A:Residues: 1-476 <ARD>
A:Cross-references: EMBL:U26430; NID:g1326068; PIDN:MAC44091.1; PID:g1326070
A:Experimental source: strain RCR2011
```

R.Ardourel, M.; Lorete, G.; Maillet, F.; Koche, P.; Truchet, G.; Prome, J.C.; Rosenberg, M.; Microbiol. 17, 687-699, 1995

A:Title: In *Rhizobium meliloti*, the operon associated with the nod box n5 comprises nodI

A:Reference number: S71357; MUID:96111489; PMID:8801423

A:Accession: S71358

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 304-339 <ARM>

A:Cross-references: EMBL:U26430

A:Experimental source: strain RCR2011 (-S047)

C:Genetics:

A:Gene: noeA

C:Superfamily: *Rhizobium meliloti* noeA protein

C:Keywords: nodulation

Query Match 57.0%; Score 45; DB 2; Length 476;
Best Local Similarity 63.6%; Pred. No. 5.7;
Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 5 QMNFAGIEAA 15
DB 100 EWSFAGLKAA 110

RESULT 3
H9313
NoeA host specific nodulation protein [imported] - *Sinorhizobium meliloti* (strain 1021)

C:Species: *Sinorhizobium meliloti*

C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001

C:Accession: H9313

R.Barrett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows

; Katman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.

Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001

A:Title: Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti*

A:Reference number: A95262; MUID:21396509; PMID:11481432

A:Accession: H9313

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-476 <NUR>

A:Cross-references: GB:AE006469; PIDN:AAK65074.1; PID:g14523508; GSPDB:GN00165

A:Experimental source: strain 1021, megaplasmid pSymA

R.Galbert, F.; Flann, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,

pele, D.; Chinn, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;

L.; Hyman, R.W.; Jones, T.

Science 293, 666-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Katman, S.; Keating, D.H.; Kiss, E.; Komp, C.; LeLaure,

hebaull, P.; Vandenberg, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.

A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation

C:Genetics:

A:Gene: noeA

C:Superfamily: *Rhizobium meliloti* noeA protein

Query Match 57.0%; Score 45; DB 2; Length 476;
Best Local Similarity 63.6%; Pred. No. 5.7;
Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 5 QMNFAGIEAA 15
DB 100 EWSFAGLKAA 110

RESULT 4
AB2491

fine protein VCA0178 [imported] - *Vibrio cholerae* (strain N16961 serogroup O1)

C:Species: *Vibrio cholerae*

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C:Accession: AB2491

R.Heldberg, J.F.; Eelsen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.

chardeon, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, H.

1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.

A:Reference number: AB2035; MUID:20406833; PMID:10952301

A:Accession: AB2491

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-217 <HEI>

A:Cross-references: GB:AE004358; GB:AE003853; NID:96657566; PIDN:AAF96091.1; GSPDB:GN

A:Experimental source: serogroup O1, strain N16961; biotype El Tor

C:Genetics:

A:Gene: VCA0178

A:Map position: 2

Query Match 54.4%; Score 43; DB 2; Length 217;
Best Local Similarity 75.0%; Pred. No. 5.5;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 TEQWNLACIHA 13
DB 170 TEQWNLACIHA 181

RESULT 5
H90825
hypothetical protein ECS1576 [imported] - *Escherichia coli* (strain O157:H7, substrain

C:Species: *Escherichia coli*

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: H90825

R.Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and g

A:Reference number: A96629; MUID:21156231; PMID:11256796

A:Accession: H90825

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-247 <HAY>

A:Cross-references: GB:BA000007; PIDN:BAK34999.1; PID:g13361040; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECS1576

Query Match 54.4%; Score 43; DB 2; Length 247;
Best Local Similarity 53.8%; Pred. No. 6.3;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 3 EQQWFGIEAA 15
DB 22 QQQWFGIEAA 34

RESULT 6
E86554
Fe-S oxidoreductase [imported] - *Chlamydia pneumoniae* (strain J138)

C:Species: *Chlamydia pneumoniae*, *Chlamydia pneumoniae*

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001

C:Accession: E86554

R.Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.

Nucleic Acids Res. 28, 2311-2314, 2000

A:Title: Comparison of whole genome sequences of *Chlamydia pneumoniae* J138.

A:Reference number: AB6491; MUID:20330349; PMID:10871362

A:Accession: E86554

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-369 <STO>

A:Cross-references: GB:BA000008; NID:96978883; PIDN:BAK98719.1; GSPDB:GN00142

A:Experimental source: strain J138

C:Genetics:

A:Gene: CP10513

C:Superfamily: hypothetical protein AF0390

Query Match 53.2%; Score 42; DB 2; Length 369;

| | | |
|-----------------------|-----------------|---------------|
| Best Local Similarity | 77.8%; | Pred. NO. 14; |
| Matches | 7; Conservative | 1; Mismatches |
| | | 1; Indels |
| | | 0; Gaps |
| | | 0; |

```
OY      6 WNFAGIEAA 14
         11: 11111
Db      305 WNYLGIEAA 313
```

RESULT 7

conserved hypothetical protein CP0240 [imported] - Chlamydomophila pneumoniae (strains CWL
N) Alternate names: fe-s oxidoreductase
C:Species: Chlamydomophila pneumoniae, Chlamydomophila pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
#accession: G72069, F81559
N:Kakmai, S.; Mitchell, R.; Marathe, R.; Lammel, C.; Pan, J.; Olinger, L.; Grimwood, J.;
N:Date Genet.: 21-385-389, 1999
N:Species: Chlamydomophila pneumoniae and C. trachomatis.
N:Accession: G72000, F82066, F82069, F82070, F82071, F82072, F82073, F82074, F82075, F82076, F82077, F82078, F82079, F82080, F82081, F82082, F82083, F82084, F82085, F82086, F82087, F82088, F82089, F82090, F82091, F82092, F82093, F82094, F82095, F82096, F82097, F82098, F82099, F82100, F82101, F82102, F82103, F82104, F82105, F82106, F82107, F82108, F82109, F82110, F82111, F82112, F82113, F82114, F82115, F82116, F82117, F82118, F82119, F82120, F82121, F82122, F82123, F82124, F82125, F82126, F82127, F82128, F82129, F82130, F82131, F82132, F82133, F82134, F82135, F82136, F82137, F82138, F82139, F82140, F82141, F82142, F82143, F82144, F82145, F82146, F82147, F82148, F82149, F82150, F82151, F82152, F82153, F82154, F82155, F82156, F82157, F82158, F82159, F82160, F82161, F82162, F82163, F82164, F82165, F82166, F82167, F82168, F82169, F82170, F82171, F82172, F82173, F82174, F82175, F82176, F82177, F82178, F82179, F82180, F82181, F82182, F82183, F82184, F82185, F82186, F82187, F82188, F82189, F82190, F82191, F82192, F82193, F82194, F82195, F82196, F82197, F82198, F82199, F82200, F82201, F82202, F82203, F82204, F82205, F82206, F82207, F82208, F82209, F82210, F82211, F82212, F82213, F82214, F82215, F82216, F82217, F82218, F82219, F82220, F82221, F82222, F82223, F82224, F82225, F82226, F82227, F82228, F82229, F82230, F82231, F82232, F82233, F82234, F82235, F82236, F82237, F82238, F82239, F82240, F82241, F82242, F82243, F82244, F82245, F82246, F82247, F82248, F82249, F82250, F82251, F82252, F82253, F82254, F82255, F82256, F82257, F82258, F82259, F82260, F82261, F82262, F82263, F82264, F82265, F82266, F82267, F82268, F82269, F82270, F82271, F82272, F82273, F82274, F82275, F82276, F82277, F82278, F82279, F82280, F82281, F82282, F82283, F82284, F82285, F82286, F82287, F82288, F82289, F82290, F82291, F82292, F82293, F82294, F82295, F82296, F82297, F82298, F82299, F82300, F82301, F82302, F82303, F82304, F82305, F82306, F82307, F82308, F82309, F82310, F82311, F82312, F82313, F82314, F82315, F82316, F82317, F82318, F82319, F82320, F82321, F82322, F82323, F82324, F82325, F82326, F82327, F82328, F82329, F82330, F82331, F82332, F82333, F82334, F82335, F82336, F82337, F82338, F82339, F82340, F82341, F82342, F82343, F82344, F82345, F82346, F82347, F82348, F82349, F82350, F82351, F82352, F82353, F82354, F82355, F82356, F82357, F82358, F82359, F82360, F82361, F82362, F82363, F82364, F82365, F82366, F82367, F82368, F82369, F82370, F82371, F82372, F82373, F82374, F82375, F82376, F82377, F82378, F82379, F82380, F82381, F82382, F82383, F82384, F82385, F82386, F82387, F82388, F82389, F82390, F82391, F82392, F82393, F82394, F82395, F82396, F82397, F82398, F82399, F82400, F82401, F82402, F82403, F82404, F82405, F82406, F82407, F82408, F82409, F82410, F82411, F82412, F82413, F82414, F82415, F82416, F82417, F82418, F82419, F82420, F82421, F82422, F82423, F82424, F82425, F82426, F82427, F82428, F82429, F82430, F82431, F82432, F82433, F82434, F82435, F82436, F82437, F82438, F82439, F82440, F82441, F82442, F82443, F82444, F82445, F82446, F82447, F82448, F82449, F82450, F82451, F82452, F82453, F82454, F82455, F82456, F82457, F82458, F82459, F82460, F82461, F82462, F82463, F82464, F82465, F82466, F82467, F82468, F82469, F82470, F82471, F82472, F82473, F82474, F82475, F82476, F82477, F82478, F82479, F82480, F82481, F82482, F82483, F82484, F82485, F82486, F82487, F82488, F82489, F82490, F82491, F82492, F82493, F82494, F82495, F82496, F82497, F82498, F82499, F82500, F82501, F82502, F82503, F82504, F82505, F82506, F82507, F82508, F82509, F82510, F82511, F82512, F82513, F82514, F82515, F82516, F82517, F82518, F82519, F82520, F82521, F82522, F82523, F82524, F82525, F82526, F82527, F82528, F82529, F82530, F82531, F82532, F82533, F82534, F82535, F82536, F82537, F82538, F82539, F82540, F82541, F82542, F82543, F82544, F82545, F82546, F82547, F82548, F82549, F82550, F82551, F82552, F82553, F82554, F82555, F82556, F82557, F82558, F82559, F82560, F82561, F82562, F82563, F82564, F82565, F82566, F82567, F82568, F82569, F82570, F82571, F82572, F82573, F82574, F82575, F82576, F82577, F82578, F82579, F82580, F82581, F82582, F82583, F82584, F82585, F82586, F82587, F82588, F82589, F82590, F82591, F82592, F82593, F82594, F82595, F82596, F82597, F82598, F82599, F82600, F82601, F82602, F82603, F82604, F82605, F82606, F82607, F82608, F82609, F82610, F82611, F82612, F82613, F82614, F82615, F82616, F82617, F82618, F82619, F82620, F826

| Query Match | Score 42: | DB 2: | Length 369: |
|-----------------------|---------------|---------------|-------------|
| Best Local Similarity | 77.88: | Pred. No. 14: | |
| Matches | 7: | Mismatches | 1: |
| | | Indels | 0: |
| | | Gaps | 0: |
| QY | 6 WNFAGIEAA | 14 | |
| | : | | |
| Db | 305 WNYLGIEAA | 313 | |

RESULT 8

glucose-6-phosphate isomerase [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
 C:Accession: E872726
 R:NIEMAN, W.C.; FELDYJUM, T.V.; PAULSEN, I.T.; NELSON, K.E.; EISEN, J.; HEIDELBERG, J.B.; LAUB, M.T.; DEBBY, R.T.; DODSON, R.J.; DUTKIN, A.S.; GWINN, M.U.; HALL, D.H.; KOLONIN, J.; ERMOLEVA, M.; WHITE, O.; SALZBERG, S.L.; SHAPIRO, L.; VENTER, J.C.; FRASER, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A:Title: Complete Genome Sequence of *Caulobacter crescentus*.
 A:Reference number: A87249; MUID:21173698; PMID:11259647
 A:Accession: E87276
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-539 <570>
 A:Cross-references: GB:AE005673; NID:g13421349; PIDN:AAK22209.1; GSPDB:GN00148
 C:Genetics:
 A:Gene: CC0222
 A:Superfamily: glucose-6-phosphate isomerase

| | | | | |
|-----------------------|-------|--------------|-------|-------------|
| Query Match | 53.2% | Score 42: | DB 2; | Length 539; |
| Best Local Similarity | 50.0% | Pred. No. | 22; | |
| Matches | 7; | Conservative | 5; | Mismatches |
| | | | 2; | Indels |
| | | | | Gaps |
| | | | | 0; |

```

OY      1 MTEQQWNEAGIEAA 14
          ::|||: |||: |||
Db      48 LSKQAWDEAGLEAA 61

```

RESULT 5

prolyl oligopeptidase family protein [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C:Accession: E87706
 R:NIEMAN, W.C.; FELDBLYUM, T.V.; PAULSEN, I.T.; NELSON, K.E.; EISEN, J.; HEIDELBERG, B.; LABB, M.T.; DEBOY, R.T.; DODSON, R.J.; DURKIN, A.S.; GWINN, M.L.; HAFT, D.H.; KOENIG, J.; ERMOLOVA, M.; WHITE, O.; SALZBERG, S.L.; SHAPIRO, L.; VENTER, J.C.; FRASER, C. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A:Title: Complete Genome Sequence of *Caulobacter crescentus*.
 A:Reference number: A87249; MUID:21173698; PMID:11259647
 A:Accession: E87706
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-723 <STO>
 A:Cross-references: GB:AE005673; MIM:q13425449; PIDN:AAK25649.1; GSPDB:GN00148
 C:Genetics: CC3687

| | | | | | | | |
|----|-----|-----------------|-----|-----------------------|--------------|---------------|---------------|
| QY | 2 | TEGQOMNFGATGMAA | 15 | 53.2%; | Score 42; | DB 2; | Length 723; |
| | : | : | | Best Local Similarity | 50.0%; | Pred. No. 30; | |
| db | 316 | TEEPMTFAGVAIPA | 329 | Matches 7; | Conservative | 2; | Mismatches 5; |
| | | | | | | | Indels 0; |
| | | | | | | | Gaps 0; |

RESULT 10

Nitrate-nitrite reductase [EC 1.6.99.1] - Escherichia coli K-12
N Alternate names: dsbB; dsbB [validated]
C Species: Escherichia coli
C Date: 10-Sep-1999 #sequence=revision 10-Sep-1999 #text-change 01-Mar-2002
C Accession: F64864; A48288; JCI109
R Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.A.: Rose, D.J.; Mau, B.; Shao, Y.
J Biol. Chem. 267, 11064-11068, 1992
A Title: The complete genome sequence of Escherichia coli K-12.
A Reference number: A64720; MUID:97426617; PMID:9278503
A Accession: F64864
A Status: nucleic acid sequence not shown; translation not shown
A Molecule type: DNA
A Residues: 1-176 <BLAT>
A Cross-references: GB:E000216; GB:U000956; NID:g1787417; PIDN:AAC74269.1; PID:g1787417
A Experimental source: strain K-12, substrain MG1655
R Misakakis, D.; Georgopoulos, C.; Raina, S.
Proc. Natl. Acad. Sci. U.S.A. 90, 7084-7088, 1993
A Title: Identification and characterization of the Escherichia coli gene dsbB, whose
A Reference number: A48288; MUID:93348217; PMID:7688471
A Accession: A48288
A Status: preliminary
A Molecule type: DNA
A Residues: 'M1',1-176 <MS>
A Cross-references: GB:I03721; NID:g398017; PIDN:AAA23711.1; PID:g398018
R Plinner, E.; Padan, E.; Schuldiner, S.
J. Biol. Chem. 267, 11064-11068, 1992
A Title: Cloning, sequencing, and expression of the nbaB gene, encoding a Na⁺/H⁺ antiporter,
A Reference number: JCI108; MUID:92285803; PMID:1317851
A Accession: JCI109
A Status: translation not shown
A Molecule type: DNA
A Residues: 'M1',1-129, 'LRASVGFPRSGNAVAARFYRLPDCRSAGG',136-137, 'PAV' <PIN>
A Cross-references: GB:883655
A Experimental source: strain K-12, substrain WJ333
A Genetics:

A:Gene: dsbb; roxB
A:Map position: 25.5 min
C:Function:
A:Description: reoxidizes dsba protein specifically
A>Note: reaction depends on the presence of oxygen
C:Superfamily: protein-disulfide oxidoreductase dsbb
C:Keywords: Oxidoreductase; redox-active disulfide; transmembrane protein
F:16-32/Domain: transmembrane #status predicted <TM1>
F:49-65/Domain: transmembrane #status predicted <TM2>
F:146-162/Domain: transmembrane #status predicted <TM3>
F:144/Disulfide bonds: redox-active #status predicted

Query Match 51.9%; Score 41; DB 1; Length 176;
Best Local Similarity 60.0%; Pred. No. 9.7;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 3 ECOMNFAGIE 12
Db 132 EROMDFLGIE 141

RESULT 11

H90838
protein-disulfide oxidoreductase ECS1680 [imported] - Escherichia coli (strain O157:H7,
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 01-Mar-2002
C:Accession: H90838
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: H90838
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-116 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA035103.1; PID:g13361144; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECS1680
C:Superfamily: protein-disulfide oxidoreductase dsbb

Query Match 51.9%; Score 41; DB 2; Length 176;
Best Local Similarity 60.0%; Pred. No. 9.7;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 3 ECOMNFAGIE 12
Db 132 EROMDFLGIE 141

RESULT 12

H85696
hypothetical protein dsbb [imported] - Escherichia coli (strain O157:H7, substrain EDL93
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 03-Jun-2002
C:Accession: H85696
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: H85696
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-116 <STO>
A:Cross-references: GB:AE005174; NID:g12514882; PIDN:AAG56036.1; GSPDB:GN00145; UMG:Z15
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: dsbb
C:Superfamily: protein-disulfide oxidoreductase dsbb

Query Match 51.9%; Score 41; DB 2; Length 176;

Best Local Similarity 60.0%; Pred. No. 9.7;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 3 ECOMNFAGIE 12
Db 132 EROMDFLGIE 141

RESULT 13

D82342
probable cadaverine/lysine antiporter CsdB VC0280 [imported] - Vibrio cholerae (strai
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: D82342
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers
I, R.R.; Mekalanos, J.D.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A>Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: D82342
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-445 <HEI>
A:Cross-references: GB:AE004116; GB:AE003852; NID:g9654687; PIDN:AAF93455.1; GSPDB:GN
C:Experimental source: serogroup O1, strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0280
A:Map position: 1
C:Superfamily: L-lysine transport protein

Query Match 51.9%; Score 41; DB 2; Length 445;
Best Local Similarity 60.0%; Pred. No. 26;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 6 WNFAGIEAFA 15
Db 199 WSEFVGEAFA 208

RESULT 14

E69435
hypothetical protein AF1486 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 08-Oct-1999
C:Accession: E69435
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
., Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kalne, B.P.; Sykes,
Smith, H.O.; Woese, C.R.; Venter, J.C.
A>Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: E69435
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-471 <ALE>
A:Cross-references: GB:AE001000; GB:AE000762; NID:g2689323; PIDN:AAB89769.1; PID:g264
C:Superfamily: Methanobacterium conserved hypothetical protein MTH943

Query Match 51.9%; Score 41; DB 2; Length 471;
Best Local Similarity 61.5%; Pred. No. 28;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 3 ECOMNFAGIEAFA 15
Db 45 ECOMDFEGKEADA 57

RESULT 15

H82256
hypothetical protein VC0978 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: H82256
R:Heideberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
Chardson, D.; Ermolaeva, M.D.; Yamathayan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F.
L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: H82256
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-38 <HEI>
A:Cross-references: GB:AE004179; GB:AE003852; NID:99655432; PIDN:AAF94140.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0978
A:Map position: 1

Query Match 50.6%; Score 40; DB 2; Length 38;
Best Local Similarity 50.0%; Pred. No. 2.8;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 MTEOQWNPAGIEAA 14
|||:|:
Db 18 MTLNQMNWIGVGA 31

Search completed: July 3, 2003, 14:07:18
Job time : 14.5 secs

LENGTH: 2293
TYPE: PRT
ORGANISM: Human
US-09-368-590-2

Query Match 48.1%; Score 38; DB 4; Length 2293;
Best Local Similarity 60.0%; Pred. No. 3.7e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 OMNFAGIEAA 14
11 :1:1111
DB 719 OMRLSGLEAA 728

RESULT 14
US-08-215-805A-80
Sequence 80, Application US/08215805A
Patent No. 5559008
GENERAL INFORMATION:

APPLICANT: Chang, Yung-Fu
TITLE OF INVENTION: LEUKOTOXIN GENE FROM PASTEURRELLA
NUMBER OF SEQUENCES: 84
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: USA

ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/215,805A
FILING DATE: 22-MAR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Timian, Susan J.

REGISTRATION NUMBER: 34,103
REFERENCE/DOCKET NUMBER: 19603/61 (D-1329A)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1636
TELEFAX: (716) 263-1600

INFORMATION FOR SEQ ID NO: 80:
SEQUENCE CHARACTERISTICS:
LENGTH: 934 amino acids
TYPE: amino acid
STRANDEDNESS:

TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Pasteurella suis
STRAIN: 5943

IMMEDIATE SOURCE:
LIBRARY: P. suis DNA in Bacteriophage lambda-dash
CLONE: (lambda)yfc33-37
US-08-215-805A-80

Query Match 46.2%; Score 36.5; DB 1; Length 934;
Best Local Similarity 50.0%; Pred. No. 2.4e+02;
Matches 8; Conservative 2; Mismatches 1; Indels 5; Gaps 1;

QY 1 MTEQOM-----NFAGI 11
11:1111111111111111
DB 463 ITQQQWNNIGNLAGI 478

RESULT 15
US-09-149-727-2
Sequence 2, Application US/09149727

Patent No. 6391547
GENERAL INFORMATION:
APPLICANT: Jefferson, Richard A.
APPLICANT: Kilian, Andrzej
APPLICANT: Keese, Paul Konrad
TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
FILE REFERENCE: 190106.405
CURRENT APPLICATION NUMBER: US/09/149,727
CURRENT FILING DATE: 1998-09-08
EARLIER APPLICATION NUMBER: US 60/058,263
EARLIER FILING DATE: 1997-09-09
NUMBER OF SEQ ID NOS: 71
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 2
LENGTH: 602
TYPE: PRT
ORGANISM: Bacillus sp.
US-09-149-727-2

Query Match 45.6%; Score 36; DB 4; Length 602;
Best Local Similarity 85.7%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 EQQWMPA 9
1111111111111111
DB 550 EQQWMPA 556

Search completed: July 3, 2003, 14:07:46
Job time: 12 secs

ATTORNEY/AGENT INFORMATION:
NAME: Bennett, Virginia C.
REGISTRATION NUMBER: 37,092
REFERENCE/DOCKET NUMBER: 627-196
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-420-2200
TELEFAX: 919-881-3175
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 410 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-579-667-2

Query Match 48.1%; Score 38; DB 1; Length 410;
Best Local Similarity 41.7%; Pred. No. 52;
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 ECOMNFAGIEA 14
:||||| : :
DB 50 KORNFGSVRSA 61

RESULT 11
US-08-579-667-6
Sequence 6, Application US/08579667
Patent No. 5705624
GENERAL INFORMATION:
APPLICANT: Fitzmaurice, Wayne P.
APPLICANT: Hellmann, Gary M.
APPLICANT: Grill, Laurence K.
APPLICANT: Kumagai, Monto H.
APPLICANT: Della-Cioppa, Guy R.
TITLE OF INVENTION: DNA SEQUENCES ENCODING ENZYMES USEFUL IN
TITLE OF INVENTION: PHYTOENE BIOSYNTHESIS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Virginia C. Bennett
STREET: 1211 East Morehead Street, PO Drawer 34009
CITY: Charlotte
STATE: No. 5705624th Carolina
COUNTRY: USA
ZIP: 28234
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/579,667
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bennett, Virginia C.
REGISTRATION NUMBER: 37,092
REFERENCE/DOCKET NUMBER: 627-196
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-420-2200
TELEFAX: 919-881-3175
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 410 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-579-667-6

Query Match 48.1%; Score 38; DB 1; Length 410;
Best Local Similarity 41.7%; Pred. No. 52;
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 ECOMNFAGIEA 14

DB 50 KORNFGSVRSA 61
:||||| : :
DB 50 KORNFGSVRSA 61

RESULT 12
US-08-579-667-8
Sequence 8, Application US/08579667
Patent No. 5705624
GENERAL INFORMATION:
APPLICANT: Fitzmaurice, Wayne P.
APPLICANT: Hellmann, Gary M.
APPLICANT: Grill, Laurence K.
APPLICANT: Kumagai, Monto H.
APPLICANT: Della-Cioppa, Guy R.
TITLE OF INVENTION: DNA SEQUENCES ENCODING ENZYMES USEFUL IN
TITLE OF INVENTION: PHYTOENE BIOSYNTHESIS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Virginia C. Bennett
STREET: 1211 East Morehead Street, PO Drawer 34009
CITY: Charlotte
STATE: No. 5705624th Carolina
COUNTRY: USA
ZIP: 28234
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/579,667
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bennett, Virginia C.
REGISTRATION NUMBER: 37,092
REFERENCE/DOCKET NUMBER: 627-196
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-420-2200
TELEFAX: 919-881-3175
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 410 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-579-667-8

Query Match 48.1%; Score 38; DB 1; Length 410;
Best Local Similarity 41.7%; Pred. No. 52;
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 ECOMNFAGIEA 14
:||||| : :
DB 50 KORNFGSVRSA 61

RESULT 13
US-09-368-590-2
Sequence 2, Application US/09368590
Patent No. 6187563
GENERAL INFORMATION:
APPLICANT: Solimena, Michele
TITLE OF INVENTION: INTERACTING POLYPEPTIDES FOR
TITLE OF INVENTION: AUTOANTIGENS OF AUTOIMMUNE DISEASES
FILE REFERENCE: 101918-200 (OCR-941)
CURRENT APPLICATION NUMBER: US/09/368,590
CURRENT FILING DATE: 1999-08-04
EARLIER APPLICATION NUMBER: 60/095,657
EARLIER FILING DATE: 1998-08-07
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-579-667-4

Query Match 53.2%; Score 42; DB 1; Length 413;
Best Local Similarity 50.0%; Pred. No. 10;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 ECOMNFAGIEA 14
|||||:::|
DB 50 ECOMNFAGSVKSA 61

RESULT 8
US-08-685-808-3
Sequence 3, Application US/08685808
Patent No. 6048715
GENERAL INFORMATION:
APPLICANT: HAYNES, CHARLES A., et al
TITLE OF INVENTION: SEPARATION AND CONCENTRATION SYSTEMS BASED
ON SOLUBLE OLIGOSACCHARIDE BINDING DOMAINS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: RAE-VENTER LAW GROUP
STREET: 260 Sheridan Ave., Ste. 440
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/685,808
FILING DATE: 24-JULY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/505,860
FILING DATE: 24-JULY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Rae-Venter, Barbara
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: CBDT.017.01US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 328-4400
TELEFAX: (650) 328-4477
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 155 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: no
US-08-685-808-3

Query Match 48.1%; Score 38; DB 3; Length 155;
Best Local Similarity 60.0%; Pred. No. 17;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 OCOMNFAGIEA 13
|||||:::|
DB 91 OCOMNFLELQA 100

RESULT 9
US-08-505-860C-3
Sequence 3, Application US/08505860C
Patent No. 6174700
GENERAL INFORMATION:
APPLICANT: HAYNES, CHARLES A., et al

TITLE OF INVENTION: SEPARATION AND CONCENTRATION SYSTEMS BASED
ON SOLUBLE OLIGOSACCHARIDE BINDING DOMAINS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: RAE-VENTER LAW GROUP
STREET: 260 Sheridan Ave., Ste. 440
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/505,860C
FILING DATE: 24-JULY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rae-Venter, Barbara
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: CBDT.017.000S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 328-4400
TELEFAX: (650) 328-4477
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 155 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: no
US-08-505-860C-3

Query Match 48.1%; Score 38; DB 4; Length 155;
Best Local Similarity 60.0%; Pred. No. 17;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 OCOMNFAGIEA 13
|||||:::|
DB 91 OCOMNFLELQA 100

RESULT 10
US-08-579-667-2
Sequence 2, Application US/08579667
Patent No. 5705624
GENERAL INFORMATION:
APPLICANT: Fitzmaurice, Wayne P.
APPLICANT: Hellmann, Gary M.
APPLICANT: Grill, Laurence K.
APPLICANT: Kumagai, Monto H.
APPLICANT: Della-Cioppa, Guy R.
TITLE OF INVENTION: DNA SEQUENCES ENCODING ENZYMES USEFUL IN
PHYTOENE BIOSYNTHESIS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Virginia C. Bennett
STREET: 1211 East Morehead Street, PO Drawer 34009
CITY: Charlotte
STATE: NO. 5705624th Carolina
COUNTRY: USA
ZIP: 28234
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/579,667
FILING DATE:
CLASSIFICATION: 435

TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 99:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-072-596-99

Query Match 100.0%; Score 79; DB 4; Length 51;
Best Local Similarity 100.0%; Pred. No. 2.7e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTEOQMFAGIEAAA 15
|||||
DB 1 MTEOQMFAGIEAAA 15

RESULT 5
US-08-465-640-2
Sequence 2, Application US/08465640
Patent No. 5955077
GENERAL INFORMATION:
APPLICANT: ANDERSEN, Peter
APPLICANT: ANDERSEN, Ase Bengaard
APPLICANT: HASLOV, Kaare
APPLICANT: SORENSEN, Anne Lund
TITLE OF INVENTION: TUBERCULOSIS VACCINE
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,640
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/123,182
FILING DATE: 20-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DK94/00273
FILING DATE: 01-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: COOPER, IVER P
REGISTRATION NUMBER: 28,005
REFERENCE/DOCKET NUMBER: ANDERSEN-3A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 95 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-465-640-2

Query Match 100.0%; Score 79; DB 2; Length 95;
Best Local Similarity 100.0%; Pred. No. 5.6e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MTEOQMFAGIEAAA 15
|||||

DB 1 MTEOQMFAGIEAAA 15

RESULT 6
US-09-001-984C-77
Sequence 77, Application US/09001984C
Patent No. 6245331
GENERAL INFORMATION:
APPLICANT: Laal, Sumar
APPLICANT: Zolla-Pazner, Susan
APPLICANT: Bellisle, John T
TITLE OF INVENTION: EARLY DETECTION OF MYCOBACTERIAL DISEASE
FILE REFERENCE: NYU-011
CURRENT APPLICATION NUMBER: US/09/001,984C
CURRENT FILING DATE: 1997-12-31
PRIOR APPLICATION NUMBER: 60/034,003
PRIOR FILING DATE: 1996-12-31
NUMBER OF SEQ ID NOS: 106
SOFTWARE: Patent Ver. 2.1
SEQ ID NO 77
LENGTH: 10
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis strain H37Rv
US-09-001-984C-77

Query Match 65.8%; Score 52; DB 4; Length 10;
Best Local Similarity 90.0%; Pred. No. 0.0024;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 TEQOWMFAGI 11
|||||
DB 1 TEQOWMFAGI 10

RESULT 7
US-08-579-667-4
Sequence 4, Application US/08579667
Patent No. 5705624
GENERAL INFORMATION:
APPLICANT: Fitzmaurice, Wayne P.
APPLICANT: Hellmann, Gary M.
APPLICANT: Grill, Laurence K.
APPLICANT: Kumagai, Monto H.
APPLICANT: Della-Cioppa, Guy R.
TITLE OF INVENTION: DNA SEQUENCES ENCODING ENZYMES USEFUL IN
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Virginia C. Bennett
STREET: 1211 East Morehead Street, PO Drawer 34009
CITY: Charlotte
STATE: No. 5705624th Carolina
COUNTRY: USA
ZIP: 28234

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/579,667
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bennett, Virginia C.
REGISTRATION NUMBER: 37,092
REFERENCE/DOCKET NUMBER: 627-196
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-881-3175
TELEFAX: 919-881-3175
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 413 amino acids

OY 1 MTEOQWNFAGIEAAA 15
|||||
DB 1 MTEOQWNFAGIEAAA 15

RESULT 2

US-08-818-111-99
Sequence 99, Application US/08818111
Patent No. 6338852

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Vardick, Thomas S.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 148
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,111
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 99:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-818-111-99

Query Match 100.0%; Score 79; DB 4; Length 51;
Best Local Similarity 100.0%; Pred. No. 2.7e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MTEOQWNFAGIEAAA 15
|||||
DB 1 MTEOQWNFAGIEAAA 15

RESULT 3

US-09-056-556-104
Sequence 104, Application US/09056556
Patent No. 6350456

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle

TREATM

STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 104:

SEQUENCE CHARACTERISTICS:
LENGTH: 51 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

US-09-056-556-104

Query Match 100.0%; Score 79; DB 4; Length 51;
Best Local Similarity 100.0%; Pred. No. 2.7e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MTEOQWNFAGIEAAA 15
|||||
DB 1 MTEOQWNFAGIEAAA 15

RESULT 4

US-09-072-596-99
Sequence 99, Application US/09072596
Patent No. 6458366

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Vardick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2003, 14:04:30 ; Search time 11 Seconds
(without alignments)
40.122 Million cell updates/sec

Title: US-09-830-839-1

Perfect score: 79

Sequence: 1 MTEQOWNFAGIEAAA 15

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 79 | 100.0 | 51 | 4 | US-08-818-112-104 |
| 2 | 79 | 100.0 | 51 | 4 | US-08-818-111-99 |
| 3 | 79 | 100.0 | 51 | 4 | US-09-056-556-104 |
| 4 | 79 | 100.0 | 51 | 4 | US-09-072-596-99 |
| 5 | 79 | 100.0 | 95 | 4 | US-08-465-640-2 |
| 6 | 52 | 65.8 | 10 | 4 | US-09-001-984C-77 |
| 7 | 42 | 53.2 | 413 | 1 | US-08-579-667-4 |
| 8 | 38 | 48.1 | 135 | 3 | US-08-685-808-3 |
| 9 | 38 | 48.1 | 155 | 4 | US-08-505-860C-3 |
| 10 | 38 | 48.1 | 410 | 1 | US-08-579-667-2 |
| 11 | 38 | 48.1 | 410 | 1 | US-08-579-667-6 |
| 12 | 38 | 48.1 | 410 | 1 | US-08-579-667-8 |
| 13 | 38 | 48.1 | 2293 | 4 | US-09-368-590-2 |
| 14 | 36 | 46.2 | 934 | 1 | US-08-215-805A-80 |
| 15 | 36 | 45.6 | 602 | 4 | US-09-149-727-8 |
| 16 | 36 | 45.6 | 607 | 4 | US-09-149-727-8 |
| 17 | 36 | 45.6 | 608 | 4 | US-08-637-670-36 |
| 18 | 36 | 45.6 | 618 | 4 | US-09-149-727-4 |
| 19 | 36 | 45.6 | 3111 | 2 | US-08-460-309-4 |
| 20 | 36 | 45.6 | 3111 | 2 | US-08-125-077-4 |
| 21 | 35.5 | 44.9 | 593 | 1 | US-08-202-389-12 |
| 22 | 35.5 | 44.9 | 593 | 1 | US-08-018-129-5 |
| 23 | 35.5 | 44.9 | 593 | 2 | US-08-448-250-5 |
| 24 | 35 | 44.3 | 360 | 4 | US-09-509-902A-9 |
| 25 | 35 | 44.3 | 600 | 6 | 5268463-2 |
| 26 | 35 | 44.3 | 602 | 2 | US-08-882-704A-5 |
| 27 | 35 | 44.3 | 602 | 4 | US-09-151-957-5 |

| | | | | | | |
|----|------|------|------|---|-------------------|--------------------|
| 28 | 35 | 44.3 | 602 | 6 | 5432081-2 | Patent No. 5432081 |
| 29 | 35 | 44.3 | 603 | 3 | US-09-149-727-6 | Sequence 6, Appl |
| 30 | 35 | 44.3 | 832 | 4 | US-08-630-820-7 | Sequence 7, Appl |
| 31 | 35 | 44.3 | 1056 | 2 | US-08-627-873-7 | Sequence 7, Appl |
| 32 | 35 | 44.3 | 1242 | 4 | US-09-488-270A-2 | Sequence 2, Appl |
| 33 | 34.5 | 43.7 | 561 | 1 | US-08-052-404-22 | Sequence 22, Appl |
| 34 | 34.5 | 43.7 | 561 | 1 | US-08-479-156-22 | Sequence 37, Appl |
| 35 | 34 | 43.0 | 67 | 1 | US-08-167-035-37 | Sequence 37, Appl |
| 36 | 34 | 43.0 | 67 | 1 | US-08-208-887A-37 | Sequence 37, Appl |
| 37 | 34 | 43.0 | 67 | 2 | US-08-539-005-37 | Sequence 37, Appl |
| 38 | 34 | 43.0 | 67 | 4 | US-09-280-598-57 | Sequence 57, Appl |
| 39 | 34 | 43.0 | 201 | 3 | US-08-816-346-6 | Sequence 6, Appl |
| 40 | 34 | 43.0 | 201 | 4 | US-09-335-411-6 | Sequence 6, Appl |
| 41 | 34 | 43.0 | 409 | 1 | US-08-190-802A-51 | Sequence 51, Appl |
| 42 | 34 | 43.0 | 409 | 2 | US-08-283-917-3 | Sequence 3, Appl |
| 43 | 34 | 43.0 | 409 | 2 | US-08-961-716-3 | Sequence 3, Appl |
| 44 | 34 | 43.0 | 409 | 4 | US-08-477-346-51 | Sequence 51, Appl |
| 45 | 34 | 43.0 | 409 | 4 | US-08-473-089-51 | Sequence 51, Appl |

ALIGNMENTS

RESULT 1
US-08-818-112-104

Sequence 104, Application US/08818112

Patent No. 6290969

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Skelly, Yasir A.W.

APPLICANT: Dillon, Davin C.

APPLICANT: Campos-Neto, Antonio

APPLICANT: Houghton, Raymond

APPLICANT: Veddyck, Thomas S.

APPLICANT: Twaizik, Daniel R.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY

NUMBER OF SEQUENCES: 153

CORRESPONDENCE ADDRESS:

ADDRESS: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/818,112

FILING DATE: 13-MAR-1997

CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-112-104

Query Match 100.0%; Score 79; DB 4; Length 51;
Best Local Similarity 100.0%; Pred. No. 2.7e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;